Funding Body

This research was funded by the Netherlands Organization for Scientific Research (NWO), as part of a Vidi grant (276-89-006) awarded to Mirjam Broersma.

International Max Planck Research School (IMPRS) for Language Sciences

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics

Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-36-3
Cover design and lay-out by Mauricio Troncoso Ruiz
Printed and bound by Ipskamp Drukkers, Enschede

The research reported in this thesis was conducted at the Centre for Language Studies, at the Radboud University in Nijmegen, the Netherlands.

Non-native phonetic accommodation in interactions with humans and with computers

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 13 juni 2022
om 10.30 uur precies

door

María de la Aurora Troncoso Ruiz
geboren op 2 november 1991
te Jerez de la Frontera, Spanje

Promotor:

prof. dr. M.T.C. Ernestus

Copromotor:

Dr. M.E. Broersma

Manuscriptcommissie:

prof. dr. J.J.M. Schoonen
prof. dr. M.G.J. Swerts (Tilburg University)
prof. dr. N. Sebastián Gallés (Universitat Pompeu Fabra , Spanje)
dr. M.C. Michel (Rijksuniversiteit Groningen)
dr. C. Cucchiarini

Table of Contents

Chapter 1

Introduction

In the process of acquiring a second language (L2), one of the most difficult skills to master is usually the pronunciation. In fact, L2 learners often speak with an accent in the L2, which contain traces of their native language (L1). For example, even highly proficient Dutch learners of English struggle with the pronunciation of some English sounds. The English consonant /θ/ in "**th**ink" is commonly pronounced as a /d/ or a /t/ (Hanulíková & Weber, 2012; Wester et al., 2007), and the vowel /æ/ in "b**a**d" is commonly produced as an /ɛ/ by this population (Elsendoorn, 1985; Wang & van Heuven, 2006). Difficulties in the pronunciation of an L2 come from differences between the sound inventories of the L1 and the L2. In the example above, the two sounds, /θ/ and /æ/, do not exist in the Dutch sound inventory. Thus, Dutch learners of English tend to use other sounds that exist in their L1 inventory instead, /d/ or /t/, and /ɛ/, respectively.

One of the ways in which L2 speakers learn to master the pronunciation of an L2 is by using the L2 in interactions. When participating in an interaction, language users can adapt the way they communicate to become more similar to their conversation partner. This phenomenon is commonly known as *accommodation* (Giles & Ogay, 2007), *alignment* (Pickering & Garrod, 2004) or *entrainment* (Brennan & Clark, 1996; Levitan & Hirschberg, 2011). Different theoretical approaches have attempted to explain this phenomenon, which has resulted in slightly different theories (Rasenberg et al., 2020). As a consequence, the terms that are used for this phenomenon have critically different connotations. *Accommodation* is understood as any adaptation speakers may apply to their speech when in an interaction (Giles & Ogay, 2007), also including the adaptations that may be triggered by previous knowledge the speaker may have about the interlocutor. For example, a speaker of a stigmatized variety of a language who interacts with a speaker of a privileged variety of the same language may accommodate to the interlocutor based on the exposure they received in the course of the conversation, but also based on previous knowledge they may have about the privileged variety (Giles & Ogay, 2007). *Alignment* and *entrainment*, in contrast, strictly refer to the speakers' imitative behavior (Pickering & Garrod, 2004). Thus, in order for a speaker to align to the interlocutor, they cannot just rely on previous knowledge but they need to be exposed to the interlocutor's speech choices. In this dissertation, we focus on accommodation.

Accommodation affects different aspects of a conversation. For example, nonverbal aspects are sometimes subject to accommodation: speakers can adapt the body posture or the gestures they use during a conversation (e.g., Chartrand & Bargh, 1999). Similarly, verbal aspects can be altered during an interaction too. Speakers have been found to adapt their choice of words (lexical accommodation, e.g., Brennan & Clark, 1996), the syntactic structures (syntactic accommodation, e.g., Branigan et al., 2000) and their pronunciation (phonetic accommodation e.g., Babel & Bulatov, 2012; Berry & Ernestus, 2018; Cohn et al., 2019; Lewandowski & Nygaard, 2018; Pardo, 2006; Pardo et al., 2010) in interactions. L2 sound learning could come via accommodation: a non-native speaker, interacting with a native speaker in the L2, could accommodate the pronunciation of specific L2 sounds, adopting the native speaker's pronunciation. If these adaptations are preserved after the interaction, accommodation could lead to the learning of L2 sounds.

Most of the literature on phonetic accommodation had been based on crossdialectal situations between native speakers of the same language (e.g., Babel, 2010; Pardo, 2006 among others). In recent decades, the scope of phonetic accommodation research has been expanding to include new situations where speakers may accommodate their speech. For example, not only has this phenomenon been documented when speakers use their native language but also when they use a second language (e.g., Berry & Ernestus, 2018; Gnevsheva et al., 2021; Hwang et al., 2015; Kim et al., 2011; Liu & Johnson, 2017, among others). Understanding the mechanisms underlying accommodation in L2 speakers could help us understand how L2 acquisition takes place in everyday conversations, outside of the language classroom settings.

Accommodation is not restricted to inter-human interactions, but it also takes place when speakers interact with computers. Due to the rapid advances in technology in recent decades, language systems that interact with humans are rather common nowadays (Branigan et al., 2010; Nass & Moon, 2000). According to the Computers As Social Actors paradigm (Nass & Moon, 2000), humans treat computers similarly to how they treat other fellow humans. Usually, this behavior is triggered by the presence of certain "human" attributes in the computers, for example, the use of speech (or words) and the interactiveness. Investigating how speakers may accommodate their pronunciation to virtual assistants helps us further understand phonetic accommodation and how L2 learning in interactions with computers may differ from L2 learning in interactions with other humans.

Considerable interest has been drawn to investigating why phonetic accommodation takes place in the first place, by disentangling which factors – social and linguistic – foster or hinder the presence of this phenomenon in interactions. This interest calls for the development of new methodologies in order to study spontaneous conversations in controlled laboratory settings, where the researcher can precisely control for the phonetic input the participant receives. With several exceptions (e.g., Berry & Ernestus, 2018; Romera & Elordieta, 2013), the study of interactions in laboratory settings tends to sacrifice the ecological validity of the conversation, i.e., the spontaneity and the naturalness of the interaction. In addition, the situations in which this phenomenon is studied are often simplified to an extent that the communicative goal is removed. For example, shadowing tasks, where a participant is asked to repeat single word utterances, are commonly used (e.g., Gnevsheva et al., 2021; Goldinger, 1998; Namy et al., 2002; Pardo, 2006). These situations, lacking the spontaneity and the communicative goal, are not directly representative of everyday interactions.

This dissertation aims to contribute to the study of phonetic accommodation from a theoretical but also a methodological perspective. First, we investigate how conversational factors affect phonetic accommodation in non-native speakers, both in inter-human interactions and interactions between humans and computers. The factors under study are (1) receiving *exposure* to the interlocutor's pronunciation, (2) receiving *feedback* implying misunderstanding from the interlocutor, and (3) the *proficiency* level in the L2 of the non-native speakers. Second, we aim to provide two suitable methodologies to study phonetic accommodation. The first methodology is an experimental paradigm that allows for the study of phonetic accommodation in controlled, yet spontaneous, interactions. The second methodology presented in this dissertation aims to optimize the resources used to evaluate the intelligibility of L2 speakers' pronunciation.

This Introductory chapter is organized as follows. Section 1.1 explains how three different factors —exposure, feedback and proficiency— may affect non-native phonetic accommodation in interactions, including human and computer interlocutors; Section 1.2 motivates the development of new methodologies to study phonetic accommodation, and Section 1.3 introduces the research questions addressed in each chapter and the methodologies used to address those questions.

1.1. What drives accommodation?

There are different mechanisms that have been proposed to explain accommodation (Branigan et al., 2010). From a sociolinguistic perspective, accommodation has been explained as a social phenomenon. Several social aspects have been identified to modulate the degree of accommodation speakers may display. For example, the power dynamics between the two participants of the conversation (Willemyns et al., 1997), the speakers' roles – passive or active (Pardo, 2006)–, their ages (Zellou et al., 2021), their genders (Babel, 2012; Pardo, 2006; Zellou et al., 2021), and the likeability of the conversation partner have been found to affect speakers' accommodating behavior. From this social perspective, a speaker may accommodate to bridge the social gap with their conversation partner (Giles & Ogay, 2007). In fact, accommodating to the interlocutor does seem to trigger more favorable reactions from conversation partners: individuals who accommodate are usually perceived as more likeable (Bradac et al., 1988), for instance.

Accommodation has also been explained from a perspective of efficiency. The majority of previous literature has followed one of two different perspectives differing in whether accommodation is proposed to be a phenomenon that facilitates the speaker's or the listener's role in the interaction. This speaker-listener debate is also at the core of other linguistic phenomena responsible for language variation, such as acoustic reduction, i.e., the process by which speakers shorten or weaken the articulation of word pronunciations (Ernestus, 2014; Turnbull, 2015).

Literature investigating the listener-driven and the speaker-driven perspectives have found that the weight of these two approaches may change depending on the specific circumstances of the situation, including whether the speaker is using their native language or second language (Costa et al., 2008) and whether they are interacting with a human or a computer (Branigan et al., 2010). The following subsections describe how accommodation is understood from the speaker-driven and the listener-driven perspectives and how these approaches are applied to contexts including non-native speakers and human-computer interactions.

1.1.1. Phonetic accommodation as an automatic mechanism

Accommodation has been argued to be an automatic mechanism that facilitates the role of the speaker in the conversation (Pickering & Garrod, 2004). Under this light, accommodation is interpreted as *priming*, i.e., the process by which exposure to a linguistic cue facilitates a speaker processing of the same cue in the future. When a speaker receives exposure to a given cue during a conversation, for example, a specific word like “cat”, they need to process this cue in order to decode the message. The higher level of activation of the cue makes it faster to be recognized or more likely to be used in the near future.

Similarly, the acoustic cues included in the speech of the conversation partner will lead to a higher degree of activation and would, therefore, be easier for the speaker to use than other cues with lower activation levels (Pickering & Garrod, 2004). As a consequence, speakers use the same cues included in their conversation partner's speech simply because it is more efficient for them to retrieve the cues to which they have already been exposed, due to the high levels of activation.

In order to confirm the validity of this perspective, studies have investigated how exposure to the speech triggers accommodation (Babel et al., 2013; Berry & Ernestus, 2018; Delvaux & Soquet, 2007; Gnevsheva et al., 2021; Goldinger, 1998; Hwang et al., 2015; Namy et al., 2002; Pardo et al., 2012). Interestingly, exposure is effective in triggering accommodation even when it is presented as passive exposure, i.e., when the speaker listens to speech but does not have actively act on it. Delvaux and Soquet (2007) found that French speakers who listened to ambient speech in a different variety of French accommodated their speech to become more similar to this variety.

In non-native contexts, exposure to the interlocutor's speech was first theorized to be less effective in triggering accommodation than in contexts involving native speakers only (Costa et al., 2008). In order to be primed by the exposure received, speakers need to have enough cognitive resources available to interpret the input and, in addition, be able to pay close attention to characteristics of the speech input they are exposed to (Costa et al., 2008). For instance, speakers should be able to attend to the acoustic characteristics of the speech of the interlocutor for phonetic accommodation to take place. Using an L2 usually requires more cognitive effort than using an L1 (Kormos, 2006; Segalowitz, 2010). Thus, speakers who use an L2 may not have enough cognitive resources available to pay attention to the characteristics of the speech and, therefore, would be less likely to be primed by the exposure to the speech of their conversation partner (Costa et al., 2008). However, evidence from studies investigating non-native phonetic accommodation suggests otherwise (Berry & Ernestus, 2018; Gnevsheva et al., 2021; Hwang et al., 2015). Even when using an L2, speakers seem to be primed by mere exposure, triggering them to accommodate their speech. For example, Hwang et al. (2015) found that Korean speakers adapted their pronunciation of English and, as a result, sounded more native-like while receiving exposure to a native speaker's pronunciation of English. The effect of exposure has also been found in more interactive settings. For instance, in Berry and Ernestus (2018), Spanish speakers who interacted with a Dutch speaker in English were found to adjust their pronunciation of a given sound contrast as a result of receiving exposure to the Dutch speaker's pronunciation of the contrast.

If the effect of exposure is solely based on automatic priming, phonetic accommodation could be expected to be triggered by exposure similarly in inter-human interactions and interactions between a human and a computer. Studies investigating the effect of exposure in these two settings, however, suggest that exposure is more effective in inducing accommodation when speakers believe the input is coming from a human interlocutor (Branigan et al., 2010; Cohn et al., 2019). Cohn et al. (2019) found that speakers who received the same type of exposure in a shadowing task accommodated more to the voices that could be identified as human than those that belonged to virtual assistants. Thus, mere exposure was found to be more effective in triggering phonetic accommodation in human-human contexts than in computer-human contexts. This finding suggests that if accommodation is an automatic mechanism, it is still mediated by social factors.

1.1.2. Phonetic accommodation as audience-design

Accommodation has also been argued to be an audience-design phenomenon (Bell, 1984). Adapting one's speech to match the conversation partner's speech could be interpreted as a way to ease the listener's effort in the course of the conversation and ensure communicative success (Bell, 1984; Clark, 1996; Pickering & Garrod, 2004). This perspective is incorporated in the Hypo-Hyperspeech theory (Lindblom, 1990), which explains variability in the speech signal as speakers' general tendency to adapt their speech to match the interlocutor. For example, adults have been found to adapt their speech when interacting with children, producing infant-directed speech, in order to facilitate children's understanding (Thiessen et al., 2005).

From the audience-design perspective, phonetic accommodation could be predicted to be boosted in non-native speakers in conversation with native speakers. In these conversations, the speakers, in general, show a greater degree of difference in pronunciation than the speakers in cross-dialectal interactions. This greater distance could lead to especially the non-native listeners having to struggle more to process the message and could be translated into a greater risk of misunderstandings taking place. Thus, non-native speakers could be expected to feel highly motivated to accommodate to the native speakers in order to bridge the distance between them and ensure the communicative success of the interaction.

The potential of misunderstandings endangering the communicative success does seem to trigger pronunciation adjustments. In L2 learning settings, corrective feedback implying problems in the comprehensibility of the L2 speaker –either implicit or explicit– seems to trigger speakers to adapt their speech (Aliaga-García & Mora, 2008; Dlaska & Krekeler, 2013; Kartushina et al., 2015; Saito & Lyster, 2012). Similarly, in conversation settings (e.g., Burnham et al., 2010) corrective feedback has also been found to trigger speakers to hyper-articulate their speech.

In human-computer settings, feedback implying misunderstanding from the computer can also trigger phonetic accommodation. In fact, direct comparisons between interactions between two humans and interactions between a human and a computer have shown that feedback implying misunderstanding is more effective in triggering phonetic adaptation when speakers think they are interacting with a computer than with a human (Burnham et al., 2010). This difference seems to show the audience-design perspective could considerably have a heavier weight in interactions between humans and computers than in interactions among humans (Branigan et al., 2010). In particular, when interacting with a computer, speakers may be more aware of the limitations of the interlocutor and, in order to maximize the success of the conversation, they may be willing to adapt their speech more than with a fellow human interlocutor.

1.1.3. The effect of proficiency on phonetic accommodation

As speakers progress in the path towards learning an L2, the way they use and process the L2 changes. For example, the L2 vocabulary size (Lemhöfer & Broersma, 2012; Miralpeix & Muñoz, 2018) and the complexity of the syntactic structures used (Ortega, 2003) increase with higher proficiency in the L2. Proficiency has also been studied as a factor possibly modulating the degree of phonetic accommodation in non-native speakers (Berry & Ernestus, 2018).

Higher-proficiency L2 speakers are known to process speech more smoothly than lower-proficiency L2 speakers. In fact, high proficiency speakers tend to process speech in the L2 more similarly to the way native-speakers do, more fluidly and with less monitoring than lower proficiency learners require (Lecumberri et al., 2010). As discussed above, from a speaker-driven perspective, accommodation is interpreted as the result of priming, which depends on the cognitive resources available for the speakers to not only process the input but also pay close attention to the characteristics of the input (Costa et al., 2008). Thus, high proficiency speakers who tend to process L2 speech more smoothly could have more cognitive resources available and, therefore, be more likely to be primed by mere exposure. In other words, high proficiency speakers could be expected to accommodate more to their conversation partner's speech than lower proficiency speakers (Berry & Ernestus, 2018; Kim et al., 2011).

Speakers with a lower level of proficiency in the L2, in contrast, tend to show a less native-like pronunciation in the L2 than higher proficiency speakers. The less native-like pronunciation can be translated into a greater distance between them and a native speaker interlocutor. If phonetic accommodation is understood as a way to ensure communicative success (Bell, 1984), low proficiency speakers, with a greater distance to a native speaker, could be predicted to be more motivated to accommodate their pronunciation in order to prevent any misunderstandings.

The effect of proficiency in L2 phonetic accommodation is currently understudied. Berry and Ernestus (2018) investigated the effect of proficiency on phonetic accommodation in human-human interactions involving Spanish and Dutch speakers, who used English as a lingua franca to communicate with each other. The results revealed that Spanish speakers with a higher proficiency level in English showed higher degrees of accommodation to the Dutch speakers than those Spanish speakers with lower proficiency levels. Thus, these results fall in line with the speaker-driven approach to explaining phonetic accommodation (Costa et al., 2008).

1.2. Studying phonetic accommodation in laboratory settings

In order to study the mechanisms underlying phonetic accommodation, researchers often create an interaction in a laboratory setting (e.g., Berry & Ernestus, 2018; Burnham et al., 2010; Romera & Elordieta, 2013). The speech of the participants of the conversation is recorded before, after, and during the interaction. The recordings are, then, analyzed in order to detect any adaptations as the result of accommodation during the interaction. The circumstances in which the interaction takes place in the lab vary drastically across studies depending on the degree of spontaneity allowed in the interaction and the degree of control over the amount and characteristics of the speech input participants of the experiment receive.

When ecological validity is prioritized, the goal of the experimental design is to allow as much spontaneity as possible in the interaction. For example, some researchers may allow participants to interact with another person while freely discussing different conversation topics (e.g., Berry & Ernestus, 2018; Romera & Elordieta, 2013). This type of settings often requires deceiving participants up to some degree in order to decrease the inhibition and the awkwardness of being recorded. For example, the participants in this type of experiments have been be recorded while they think they are waiting for the experiment to start (Kouwenhoven et al., 2018) or for a microphone to be fixed (Torreira & Ernestus, 2012). This level of deception contributes to creating ecologically alid settings and facilitates the generalization of the participants' behavior during the experimental session to everyday interactions. However, this level of spontaneity also means that the researchers have no control over relevant aspects of the conversation, such as the content or the exact realization of the words, which may affect the degree of phonetic accommodation participants show.

In order to exercise some degree of control over a spontaneous interaction, a confederate, i.e., a person who is instructed on how to behave during the experimental session, has been included in the conversation (e.g., Kouwenhoven et al., 2018; Romera & Elordieta, 2013). Including a confederate helps homogenizing the experimental sessions: all participants in the experiment are exposed to the same conversation partner and the topics of conversation are restricted. However, a confederate does not always guarantee complete homogeneity across experimental sessions (Lewandowski & Nygaard, 2018). For instance, the confederate's speech –including the amount and the acoustic characteristics of their speech, or the degree of accommodation the confederate shows to participants– cannot be controlled to be exactly the same across participants, even when the confederate receives explicit instructions.

In the field of syntactic or lexical accommodation, the speech of the confederate can be relatively easily controlled in spontaneous conversations. For example, numerous studies have made use of the Confederate-Scripting paradigm (Branigan et al., 2000). With this paradigm, participants believe they are interacting with a confederate spontaneously but, in reality, the confederate's contribution to the conversation has been carefully scripted. Scripting the confederate's speech allows for control over the timing and the amount of exposure participants receive to a specific syntactic structure or a lexical item. However, when studying phonetic accommodation, a more exhaustive control over the input is needed in order to ensure the amount and of input and its characteristics are identical across participants.

This fine-grained level of detail over the confederate's speech can be achieved by using pre-recorded speech (Eijk et al., 2019; Gessinger et al., 2019; Gnevsheva et al., 2021; Goldinger, 1998; Namy et al., 2002; Pardo, 2006; Pardo et al., 2017). For example, in Gessinger et al. (2019), pre-recorded speech was used to ensure that participants of the experiment were equally exposed to a specific dialectal variant of the pronunciation of a German consonant. The implementation of pre-recorded speech is typically done in computer-based settings, such as the Wizard of Oz (Riek, 2012). In this paradigm, participants think they are interacting with a "smart" computer but, in reality, the computer is managed by a human. In this case, the implementation of pre-recorded speech directly leads to the dehumanization of the confederate. The use of a computer, instead of a human confederate, is claimed to affect the degree of phonetic accommodation displayed by participants (as discussed above; Branigan et al., 2010).

In addition to the dehumanization of the confederate, the use of computer-based tasks often leads to a simplification of the conversation. For example, shadowing tasks, where a speaker is asked to repeat, or to shadow, a word produced by another speaker, have been used to study phonetic accommodation (e.g., Gnevsheva et al., 2021; Goldinger, 1998; Namy et al., 2002; Pardo, 2006; Pardo et al., 2017). This type of task allows researchers to study this phenomenon stripping away social or conversational aspects, and ensuring that participants produce certain sounds in specific wordcontexts (Pardo, 2006). However, the use of oversimplified tasks, where speakers may even be indirectly encouraged to imitate their conversation partner (Pardo, 2006), and the dehumanization of the confederate hinder the generalization of the patterns found in this type of experimental sessions to everyday interactions. The study of phonetic accommodation therefore would benefit of a new experimental methodology that enables researchers to implement pre-recorded speech, unbeknownst to participants, in spontaneous interactions.

1.3. Outline and research questions

This dissertation aims to expand the knowledge on non-native phonetic accommodation. The goal is to better understand the underlying mechanisms of L2 sound learning in everyday interactions. It does so by disentangling the factors that drive phonetic accommodation in non-native speakers in four experimental chapters, where Chapter 2 develops the methodology of Chapters 3, 4 and 5.

The studies described in Chapters 3, 4 and 5 focus on Dutch learners of English with a proficiency level of English ranging from B1 to C2 – according to the Common European Framework of reference for languages (Council of Europe, 2001). This population of learners has been reported to struggle with the English vowel contrast in “bad”-“bed”, which they tend to merge into an in-between sound (Elsendoorn, 1985; Wang & van Heuven, 2006). Therefore, the Dutch participants were expected to accommodate to a native speaker of English by adapting their pronunciation of these two vowels to make them more distinct.

Chapter 2 presents a new experimental paradigm to study interactions in laboratory settings, the Ventriloquist paradigm. With the Ventriloquist paradigm, participants engage in a conversation with a human confederate who is present in the same room. Both the participant and the confederate are equipped with headphones and microphones. The participants are told they will listen to the confederate through the headphones but, in reality, they listen to pre-recorded audio samples, which the confederate plays by pressing buttons on a numeric keyboard hidden from the participant’s sight. This paradigm aims to allow researchers to implement fine-grained phonetic control over the input participants receive in ecologically valid situations. The chapter provides a detailed description of the implementation of this paradigm and evaluates the credibility and validity of this paradigm by answering two research questions. First, is the Ventriloquist Paradigm convincing enough to lead naïve participants to believe they are having a live conversation with a human confederate? Second, does this paradigm elicit more spontaneous and interactive conversations than a similar setting where the confederate is presented as a computer?

Chapter 3 investigates whether speakers accommodate differently when they interact with a human than with a computer. This chapter aims to answer the following questions: first, do speakers align their pronunciation differently to humans and to computers when the settings are identical? Second, does exposure to the interlocutor's pronunciation and/or corrective feedback on the speaker's pronunciation trigger phonetic accommodation to humans and to computer differently? In order to answer these questions, this chapter presents a direct comparison between phonetic accommodation in two types of interaction: a *human-human* interaction, where speakers believe they are interacting with a human confederate sitting in the same room (using the Ventriloquist paradigm), and a *human-computer* interaction, where participants believe they are interacting with a computer, in the style of the Wizard of Oz (Riek, 2012). Participants in this study, Dutch learners of English, either interacted with a confederate who was presented as a native speaker of English or with a "smart" computer. The Dutch learners' pronunciation of the difficult English vowel contrast in "bad"-"bed" was analyzed acoustically in order to determine whether participants adapted their pronunciation from before to after the interaction. In these two settings, the presence of exposure to the native speaker's pronunciation of the vowel contrast as well as corrective feedback on the speaker's pronunciation of this same contrast were contrasted. The data collected in Chapter 3 is used in Chapter 4 and 5 to answer further research questions.

Previous literature on the assessment of phonetic accommodation has argued the importance of using both acoustic analyses and perceptual measures to assess the degree of accommodation in terms of pronunciation. In Chapter 4, three different perceptual measures are used in order to investigate whether non-native speakers' accommodation, triggered by exposure and feedback could be perceived by the interlocutor. Native listeners of English were asked to listen to word-tokens from the speech collected in Chapter 3. They were asked to identify (perceptual measure 1) the words produced by the Dutch learners before and after their interaction with the native speaker or the computer, and to rate how native-like their pronunciation sounded (perceptual measure 2). In addition, this chapter presents the second methodological contribution of the thesis. This chapter attempts to validate the replacement of the intelligibility perceived by human listeners with artificial judgements. Here, we present a comparison between the measure of the intelligibility based on human listeners responses (perceptual measure 1) and on the performance of an Automatic Speech Aligner (perceptual measure 3).

Chapter 5 aims to contribute to a better understanding of the role of the proficiency in non-native phonetic accommodation. The study described in this chapter reanalyzes the data from Chapter 3 comparing the phonetic accommodation from nonnative speakers with high and low proficiency. It investigates the effect of proficiency on phonetic accommodation taking into account the type of interlocutor – human or computer – and the presence of exposure and feedback.

Finally, Chapter 6 presents the findings of each of the chapters included in this dissertation and discusses, based on these findings, general conclusions about the underlying mechanisms of phonetic accommodation in non-native speakers, both in inter-human interactions and interactions between humans and computers. In addition, this chapter also includes recommendations for the methodology used to study phonetic accommodation and for L2 language teaching. Furthermore, this chapter also includes suggestions for further research that follow from the findings described in this dissertation.

Chapter 2

The ventriloquist paradigm:

Studying speech processing in conversation with experimental control over phonetic input

This chapter is based on:

Felker, E., Troncoso-Ruiz, A., Ernestus, M. & Broersma, M. (2018). The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America, 144*(4), EL304–309.

ABSTRACT

This chapter presents the ventriloquist paradigm, a novel method for studying speech processing in dialogue whereby participants interact face-to-face with a confederate who, unbeknownst to them, communicates by playing pre-recorded speech. Results show that the paradigm convinces more participants that the speech is live than a setup without the face-to-face element, and it elicits more interactive conversation than a setup in which participants believe their partner is a computer. By reconciling the ecological validity of a conversational context with full experimental control over phonetic exposure, the paradigm offers a wealth of new possibilities for studying speech processing in interaction.

2.1. Introduction

This chapter presents a novel experimental paradigm that, for the first time, enables the study of speech processing in interaction while maintaining full experimental control over phonetic exposure. Speech perception and production are doubtlessly shaped by experiences in conversation, as demonstrated by research on perceptual adaptation (e.g., Norris et al., 2003) and phonetic alignment and accommodation (e.g., Pardo, 2006). To reach a fuller understanding of the mechanisms underlying language processing in interactive contexts, researchers have called for studying language perception and production in more contextualized, ecologically valid settings, such as informal face-to-face communication centered on joint tasks (e.g., Tanenhaus & Brown-Schmidt, 2008; Tucker & Ernestus, 2016; Willems, 2015). For experiments investigating the underlying mechanisms of perceptual learning and phonetic alignment, in which the quantity, context, and timing of exposure to critical speech sounds are theorized to play a key role, control of phonetic detail is crucial. However, controlling phonetic input in a natural conversation poses a methodological challenge.

All approaches to studying sound learning and adaptation make trade-offs between ecological validity and experimental control. Traditional phonetics experiments that control the type and presentation of stimuli (e.g., categorization, discrimination, shadowing, lexical decision, and judgment) have led to fundamental insights into how speech processing works in individuals when tested in isolation but do not address naturalistic interaction. Other research methods provide more ecological validity (e.g., Map task, Anderson et al., 1984; Diapix task, Van Engen et al., 2010; spontaneous dialogue, Torreira & Ernestus, 2010; Pardo et al., 2012) but do not control the phonetic exposure participants receive.

To study *syntactic* alignment, the ‘confederate-scripting’ paradigm (Branigan et al.,2000) combines natural interaction with experimental control of language input by fully scripting the linguistic input at the syntactic and lexical level. To investigate *sound* learning mechanisms, however, the relevant level to control is phonetics. Whereas phonetic studies often involve artificial accents, manipulated speech sounds, or avoidance of specific sounds, even a phonetically trained confederate cannot perfectly control all the phonetic details of their speech during a live experiment. Furthermore, since subtle phonetic alignment often occurs in dialogue (e.g., Pardo, 2006), the confederate's accent risks converging toward that of participants, such that not all of them receive comparable phonetic input. In fact, variability in the speech input can only be avoided if the speech is pre-recorded.

We introduce the new ventriloquist paradigm, which solves the problem of variable phonetic input in live speech by employing pre-recorded speech covertly in a real-time conversation. In this paradigm, a participant and confederate work together face-to-face on a cooperative computer-based task. While the participant believes they are having a normal conversation, the confederate does not actually speak but plays pre-recorded utterances to the participant's headphones while briefly hiding her face behind a screen. As in a ventriloquist performance, the true source of the confederate's speech is thus disguised. The pre-recorded speech meets the experiment's phonetic requirements and includes all phrases necessary for the joint task and various other phrases to respond to whatever the participant says.

This chapter presents the methodology of the ventriloquist paradigm and the steps required to incorporate pre-recorded speech in an experiment while convincing participants they are having a live conversation. To illustrate how the paradigm can be used to study sound learning in speech perception and production, we describe its implementation in two dialogue elicitation tasks and an auditory lexical decision test. We also evaluate the ventriloquist paradigm's effectiveness and compare it to two control setups that vary in how present or personal the confederate is: In one version, we removed the face-to-face aspect of the interaction by putting the participant and confederate in separate testing booths. In another, we further reduced the "human" nature of the interaction by not only having participants alone in a booth but also telling them they were interacting with a computer, thus implementing a 'Wizard of Oz' experiment (Fraser & Gilbert, 1991; Riek, 2012). By analyzing the conversational interaction produced with the ventriloquist paradigm and these control methods, we assess how effective the ventriloquist paradigm is at creating a convincing, interactive dialogue.

2.2. Ventriloquist paradigm methodology

2.2.1. General procedure

At the beginning of a session, the participant is told that he will play a cooperative computer game with a partner. The experiment leader explains that both players will speak into microphones and that their speech will be transmitted to each other's noise-cancelling headphones, which they must keep on throughout the session. To prevent the participant from engaging with the confederate before she can play her pre-recorded speech, the experiment leader holds the conversational floor so that the players cannot speak to each other until their headphones are on.

During the cooperative game, the participant and confederate sit at a table across from each other, each facing their own computer monitor, but with ample room between the monitors for them to see each other. Every time the confederate needs to speak, she leans toward a dummy microphone next to the table, thereby hiding her entire face behind her monitor, and surreptitiously presses a key on a hidden numeric keypad corresponding to a desired speech function. The computer then plays a pre-recorded utterance, which the participant hears in his headphones.

2.2.2. Software and speech materials

The experiment software implements a structured, collaborative two-player game that requires the players to communicate orally to share information or give each other instructions. Each key of the numeric keypad is mapped to a different audio category so that when it is pressed, an audio file from the associated speech category is played. A visual reference of the number key-audio category mappings is overlaid on the confederate's screen as a memory aid. The audio files consist of various categories of pre-recorded utterances that are scripted to meet the researcher's desired phonetic constraints. The utterances can be one of two types: trial-linked or flexible.

Trial-linked utterances can only be played on specific trials or time points within the experiment. For instance, a recording of the speaker introducing herself may be linked to the welcome screen and a recording of her saying goodbye to the end screen. Most trial-linked utterances relate to visual stimuli that occur on specific trials, such as descriptions of a displayed picture or instructions for the participant to click on a displayed word. In case participants ask the confederate to repeat herself, trial-linked utterances have follow-up versions that can be played in succession if necessary. For example, if the first utterance for a trial is "Now we want the word *flower*", a follow-up version could be "I said *flower*", and a second follow-up could be "*Flower*" with even more emphasis. The phrases vary in structure and wording to avoid repetitiveness and contain some disfluencies to make them sound more natural, but they are nevertheless kept short to reduce the chance of the participant interrupting them. To facilitate the confederate's task of playing the audio files, the software links all trial-linked utterances to a single numeric key, and pressing that key will play only the utterances linked to the current trial, in the pre-specified order.

Other pre-recorded utterances are flexible, meaning they are playable throughout the experiment to respond to whatever the participant might ask. Important flexible utterance categories include affirmative responses, negative responses, backchannels such as "mm-hm", variations of "I don't know" (also useful for responding to off-topic remarks or open-ended questions), requests to elaborate, reassuring remarks, thank-yous, utterances of surprisal about the appearance of new trials (if the confederate cuts a trial short to unblock the conversation), and reminders of the task rules. Each category contains numerous recordings that serve the same communicative function, and there are enough utterances to ensure that no audio file is repeated within a session.

2.2.3. Physical setup and equipment

The ventriloquist paradigm is set up in a large booth or testing room, ideally with a window through which the experiment leader can monitor the activity. A single computer runs the experiment software and displays graphics on two wide monitors situated side by side, facing opposite directions across the table. A numeric keypad with silent keys is just below the table (e.g., resting on a cabinet), hidden from the participant's view. At the center of the table rests an active microphone aimed toward the participant and connected to an audio mixing console. The confederate's dummy microphone stands at the outside edge of the confederate's side of the table.

Audio output from the computer is split into two channels: one to the participant's noise-cancelling over-ear headphones, and one to the audio mixing console. The console combines audio input from the computer and participant's active microphone and sends it to the confederate's headphones, an audio recorder, and a pair of headphones outside the testing booth for the experiment leader.

2.3. Examples of Ventriloquist Paradigm Implementation

To illustrate how the ventriloquist paradigm can be used to answer specific research questions about speech perception or production in interaction, this section presents two dialogue elicitation tasks and an auditory lexical decision task we have implemented with it.

2.3.1. Dialogue elicitation task: Code Breaker game

The Code Breaker game is designed for research into various types of phonetic learning, such as perceptually adapting to an unfamiliar accent's vowel shift or learning to more clearly produce a difficult non-native sound contrast. While critical speech sounds in the ventriloquist's speech repertoire are controlled to provide the desired type and amount of phonetic input for participants to learn from, various task- and interaction- related variables, such as the presence and type of feedback from the confederate, can also be manipulated to test specific hypotheses about learning mechanisms.

In the Code Breaker game, the participant and confederate work together to solve puzzles and tell each other to click on words belonging to phonological minimal pairs, with or without feedback. In each trial (Figure 1a), Player A sees a sequence of colored shapes followowed by a question mark, above an array of four words, and he must tell his partner what shape comes next. Player B finds the specified shape on her screen and tells her partner to click on the target word linked to that shape. When the ventriloquist is Player A, trial-linked utterances refer to a puzzle's solution (e.g., "I think we need a black square"); when she is Player B, the trial-linked utterances contain the target words (e.g., "So you should click on *land*"). For the study of speech perception, the participant acts as Player A, as their challenge is to accurately perceive the target words. For production, the participant acts as Player B, as their challenge is to pronounce the target words accurately.

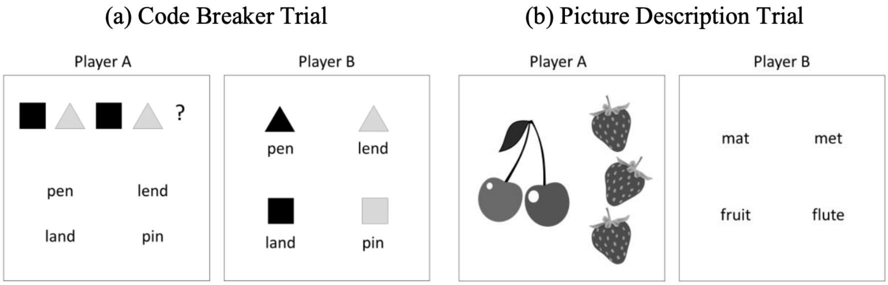


Figure 1. Sample screens for two players in one trial of the Code Breaker game (Section 2.3.1) and for one trial of the picture description game (Section 2.3.2).

2.3.2. Dialogue elicitation task: Picture description

Another interactive game involving more elaborate and contextualized speech is the picture description task (Figure 1b), which can be used in combination with Code Breaker trials to give the participant different types of phonetic exposure (e.g., hearing words in various semantic contexts, with or without their phonological neighbors, with or without spelling cues, etc.). In each picture description trial, Player A sees a picture while Player B sees an array of four words consisting of two phonological minimal pairs. Player A describes the picture until Player B is able to select the word matching the described picture. Optionally, Player B is also instructed to read aloud their four word options before making a final choice. If the ventriloquist is Player A, the trial-linked utterances are the picture descriptions; if she is Player B, they are the speaker declaring her answer (e.g., "I have mat, met, fruit, and flute, so I'm going to choose fruit").

2.3.3. Auditory lexical decision task

An auditory lexical decision task can be employed to measure the participant's perceptual adaptation to the pre-recorded speaker after a dialogue elicitation task. This method is identical to a regular lexical decision test except the participant believes they are responding to words being read aloud in real-time by their conversation partner. The participant is instructed not to request repeats or clarification to ensure that he does not try to interact during this test, and the confederate remains hidden behind her monitor the entire time to avoid visual distraction. The trial-linked audio consists of the auditory lexical decision stimuli. Rather than being triggered by the confederate's button presses, it is played automatically at pre-determined inter-stimulus intervals, randomized within a small range to give the impression that the items are being read in real time.

2.4. Validity of the ventriloquist paradigm

The validity of the ventriloquist paradigm depends on how reliably it convinces participants they are engaged in a genuine conversation. We analyze the participant-ventriloquist interaction using data from 101 Dutch participants (aged 18-30 years) speaking their highly proficient L2 English in sessions of 15 to 30 minutes in one of two experiments, each with a different confederate and different pre-recorded native English speaker. One experiment (56 participants) used the Code Breaker (production) and picture description task, and the other (45 participants) used the Code Breaker (perception) and lexical decision task.

All participants engaged with the cooperative tasks, and nobody overtly questioned the genuineness of the conversations during the session. Questionnaires administered at the end of each session, without the confederate present, showed that 79.2% of participants reported no suspicion that their partner's speech was prerecorded. The most common reasons given for suspecting pre-recorded speech were that the timing of the ventriloquist's speech or body movements felt slightly off, phrase structures were repeated, or the speech sounded "too perfect."

Interestingly, we found two differences between those who did and did not report suspicions. For the former group, interactivity (as measured by the total number of ventriloquist utterances played during the entire Code Breaker game) was lower than for the latter (mean = 88.5 utterances vs. mean = 96.6 utterances, $t(42.075) = 2.20$, $p = 0.03$). This suggests either that hearing more ventriloquist speech increased believability, or, alternatively, that participants sought less interaction when they suspected their partner's speech was not live. Furthermore, self-reported English proficiency (speaking, listening, reading, and writing) was higher for those who did report suspicions than for those who did not ($t(33.56) = 2.28$, $p = 0.03$), suggesting either that greater task difficulty increased participants' susceptibility to the illusion or that discovering the truth increased self-ratings. Between the two experiments, the proportion of participants who bought into the illusion did not differ; $\chi2$ $(1, N = 101) = 0.50$, $p = 0.48$.

2.5. Evaluating the importance of face-to-face context

To determine whether the face-to-face setting of the ventriloquist paradigm affected the extent to which participants believed in the genuineness of the conversation, we collected data from 22 new participants from the same population using an alternative setup in which the participant and confederate did the same tasks together but in separate testing booths from which they could not see each other.

In these experiments, importantly, the tasks, confederates, audio setup, software, and pre-recorded speech materials were the same as in Section 2.4, except no dummy microphone was needed since the participant never saw the inside of the confederate's booth. For the interactive games, the confederate used her keyboard to play prerecorded speech exactly as in the ventriloquist paradigm, aiming to be just as interactive as in the ventriloquist setup to enable a fair comparison. While the confederate never entered the participants' testing booth, participants could see her walking by the window of their booth and heard the experiment leader speaking to her as if she was another participant. Furthermore, whenever the experiment leader gave instructions to the participant, she then stopped inside the confederate's booth to create the impression that she instructed her as well.

In the post-experiment questionnaire, only 32% of the participants reported no suspicion that the speech was pre-recorded, a significantly lower proportion than in the ventriloquist paradigm ($\chi2(1, N = 123) = 17.375, p < 0.001$); moreover, all the participants who noticed the pre-recorded speech also believed their partner was actually a computer or robot. These results suggest that the face-to-face aspect of the ventriloquist paradigm strongly contributed to making the pre-recorded speech sound live. The separate-booth setup, on the other hand, does not seem viable for studying natural conversation, as it convinced few participants that they were having a live conversation or were even talking to another human.

2.6. Evaluating the importance of beliefs about interlocutor's humanness

To examine whether the ventriloquist paradigm creates more engaging and interactive conversation than when people believe they are talking to a computer, 36 additional participants from the same population were tested in a new setup in which they were told upfront that they were interacting with a computer.

The setup and procedure were as described in Section 2.5, except that the experiment leader told participants that their partner was a smart computer player and no attempt was made to hide the fact that the speech was pre-recorded. Participants completed the tasks from the second experiment (Code Breaker perception and lexical decision task) described in Section 2.4, and the pre-recorded speech was played by the same person as before, who again aimed to be just as interactive as in the ventriloquist setup.

Nobody reported any suspicion that they had been playing with a real person rather than with a smart computer. We compared the interactions during the computer player version of the Code Breaker game to those from the matching ventriloquist paradigm sessions. The total number of pre-recorded utterances played per session was similar in the ventriloquist setup (mean = 112.1, SD = 15.8) and the computer-player setup (mean = 120.1, SD = 21.1), ($t(63.3) = 1.904$, $p = 0.06$), confirming that the computer player and ventriloquist were played in a comparable way. To assess the interactivity of the conversation, we measured how often and for how long participants spoke, excluding two participants due to recording malfunctions. The *number* of participant utterances (utterances being defined as any stretches of speech bounded by either a pause of at least 0.6 seconds or an intervening pre-recorded utterance) was higher in the ventriloquist setup (mean = 178.8, SD = 50.8) than in the computer-player setup (mean = 136.6, SD = 41.6); $t(76.46) = 4.06$, $p < 0.001$. For participants' speech duration, we analyzed the ratio of participant-to-confederate speaking time, rather than participant speaking time alone, to control for the influence of any between-session variability in confederate speech duration. This ratio was significantly higher in the ventriloquist setup (2.05:1) than in the computer-player setup (1.76:1); $t(58.94) = 2.05$, $p = 0.04$. These results demonstrate that the ventriloquist paradigm increases participants' engagement in the conversation, as measured by their speaking behavior, relative to the computer-player control setup.

2.7. General discussion

This chapter described the ventriloquist paradigm, a novel experimental method that incorporates pre-recorded speech in real-time, face-to-face conversation. The results showed that the ventriloquist paradigm convinces most participants that they are having a genuine dialogue. The face-to-face aspect of the interaction appears to be instrumental in maintaining the illusion, as participants were much less likely to notice that the speech was pre-recorded in the ventriloquist paradigm than in a control setup utilizing separate testing booths. Participants may assume, possibly based on prior experience with experiments, that the speech they hear from headphones in a testing booth is pre-recorded unless they have strong evidence to the contrary, such as the confederate's physical co-presence. Furthermore, analyses showed that the ventriloquist paradigm elicited more interactive, engaging conversation than a setup in which participants believed they were interacting with a computer.

Practical challenges associated with the ventriloquist paradigm are that scripting and recording the ventriloquist's utterances is time-consuming, and the paradigm requires a confederate with some degree of acting ability who can think on her feet. Moreover, researchers might have to discard some data from participants who did not buy into the ventriloquist illusion. Furthermore, compared to ordinary conversation, the spontaneity and complexity of interaction with the ventriloquist will always be somewhat limited, given that the pre-recorded speech is only designed to handle conversation around highly-structured, predictable tasks. However, we believe that the paradigm can be adapted to incorporate more complex dialogue tasks than we have used so far, such as the Map Task (Anderson et al., 1984), although extensive pilot testing would be needed to determine what trial-linked and flexible utterances would be necessary to make the interaction convincing. Finally, it should be noted that using pre-recorded speech precludes any level of linguistic alignment from the confederate to the participant, and this lack of reciprocal alignment, while enabling full control over the phonetic characteristics of the input, necessarily makes the interaction less natural than if the confederate were speaking spontaneously.

In short, the ventriloquist paradigm can be used to study how people learn from and adapt to each other's speech in everyday communication centered on cooperative tasks, which affords more ecological validity than many traditional experimental paradigms. As the paradigm can be used with a variety of different cooperative tasks, numerous task- and interaction-related variables can be manipulated to study various aspects of speech perception and production. Most importantly, the ventriloquist paradigm allows researchers to fully control the phonetic input participants receive in the conversation, thereby facilitating research into the underlying mechanisms of sound learning. By combining this fine-grained control of the input with a naturalistic dialogue, the ventriloquist paradigm opens up a wealth of new possibilities for studying speech processing in interaction.

Chapter 3

Comparing phonetic accommodation to humans and to computers in interactive settings:

effects of exposure and feedback

This chapter is based on:

Troncoso-Ruiz, A., Ernestus, M. & Broersma, M. (2019). Learning to produce difficult L2 vowels: The effects of awareness-raising, exposure and feedback. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th ICPhS, Melbourne, Australia 2019* (pp. 1094–1098). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

ABSTRACT

This chapter presents a comparison between human-human and human-computer phonetic accommodation in an interactive context. The two settings only differ in whether the interlocutor was presented as a computer or as a human. Neither of the interlocutors was speaking live during the interaction, pre-recorded speech was used unbeknownst to participants in both settings to ensure optimal control over the phonetic input participants received. This illusion was maintained by using the Ventriloquist Paradigm (Chapter 2). Accommodation was measured as the degree to which Dutch speakers produced more native-like tokens of the English /æ-ɛ/ vowels after interacting in English with a computer or a human confederate. During the interaction, the exposure to the native pronunciation of the vowels (presence/absence) and feedback on the non-native pronunciation (presence/absence) were fully crossed, resulting in four between-participant conditions in each setting. Acoustic analyses show that exposure to the vowels and feedback fostered more phonetic accommodation to the human interlocutor than to the computer. However, the combination of exposure to the vowels and feedback triggered more accommodation in the computer setting than in the human setting. This study shows that, even when speakers perform the same task and are exposed to the same speech input, they accommodate differently to humans than to computers. Our results raise questions about the generalizability of using computer-based approaches to study human-human interactions in general and phonetic adaptation in particular.

3.1. Introduction

In conversations, people tend to adapt their pronunciation to match the speech of the interlocutor (Brennan & Clark, 1996; Pickering & Garrod, 2004; Pardo, 2006; Giles & Ogay, 2007; Levitan & Hirschberg, 2011) and they do so differently when they interact with another human being than when they interact with a computer (e.g., Branigan et al., 2010; Cohn et al., 2019; Raveh, Steiner, et al., 2019). Studying this phenomenon unavoidably entails methodological choices about the balance between phonetic control and the level of interactivity of the conversational setting. Previous studies have tended to prioritize phonetic control over interactivity by, for instance, using pre-recorded speech, even when this leads to sacrificing the spontaneity and interactivity of the conversation. The current chapter takes a different approach and aims to study the mechanisms underlying phonetic accommodation to humans and to computers while combining both phonetic control and spontaneity in the interaction. In order to do so, we have used the Ventriloquist paradigm (Chapter 2), a methodology which allows for the implementation of pre-recorded speech in an interaction unbeknownst to participants, who believe they are having a spontaneous interaction. In particular, we examine how the degree of accommodation to other humans and to computers is affected by the presence of exposure to the interlocutor's pronunciation of a specific vowel contrast and by the presence of feedback on the intelligibility of the speaker's pronunciation on the same contrast.

People have been shown to adapt their speech to the interlocutor in various ways. One is by making their speech more similar to that of the interlocutor (Branigan et al., 2010; Hwang et al., 2015). This phenomenon is known as *accommodation* (Giles & Ogay, 2007), *alignment* (Pickering & Garrod, 2004) or *entrainment* (Brennan & Clark, 1996; Levitan & Hirschberg, 2011). These different terms come with different connotations. Accommodation is used to describe speakers' adaptive behavior during an interaction, based on current exposure or previous knowledge. Alignment is generally used to describe repetitive behavior resulting from receiving explicit exposure during a conversation. Here, we study phonetic adaptation understood as accommodation.

People have also been shown to accommodate by making their speech more dissimilar to the interlocutor's speech (Babel, 2010), or more intelligible for the interlocutor (Burnham et al., 2010; Hwang et al., 2015). Accommodation has been reported to take place at many linguistic levels, including the phonetic. Speakers have been shown to adjust supra-segmental features, including their pitch (Babel & Bulatov, 2012; Gijssels et al., 2016; Levitan & Hirschberg, 2011; Pardo et al., 2013), intonation contours (Gessinger et al., 2019) speech rate (Levitan & Hirschberg, 2011), and the pronunciation of specific sounds (Babel, 2012; Delvaux & Soquet, 2007; Pardo, 2006; Pardo et al., 2010, 2012). In this chapter, we focus on this last type of phonetic accommodation regarding the accommodation of individual segments.

The majority of studies investigating how speakers can adjust this wide range of phonetic cues in interaction usually focuses on conversations including speakers who speak different dialects of the same native language (L1) (Babel 2010, 2012; Pardo et al. 2012; Willemyns et al. 1997, among many others), but it has also been shown for interactions involving one or more speakers using their second language (L2) (Berry & Ernestus, 2018; Costa et al., 2008; Hwang et al., 2015; Lewandowski & Nygaard, 2018). Conversations including non-native speakers offer a suitable context for the study of phonetic accommodation because the differences between the interlocutors' pronunciation is usually greater than when speakers share the same L1; therefore, they offer ample room for adaptation (Kim et al., 2011). In addition, studying phonetic accommodation in non-native speakers can help us understand the role of interactions with native speakers in the process of L2 acquisition. The current chapter aims to contribute to the literature on non-native phonetic accommodation by investigating the accommodation displayed by native speakers of Dutch when they interact in their L2 English.

3.1.1. The challenge of studying phonetic accommodation in laboratory settings

In order to investigate interactions, researchers often opt for laboratory settings. In these settings, a real-life conversation can be recreated under experimental control by asking participants to take part in a conversation or interactive task with an interlocutor, who can be another participant or a confederate. Both human (e.g., Berry & Ernestus 2018; Braningan et al. 2010) and computer (e.g., Babel et al., 2013; Eijk et al., 2019) confederates have been used. The two options come with advantages and disadvantages.

Human confederates have the advantage of allowing for a high level of spontaneity during the interaction, and elicit a larger degree of interactivity from participants than when they interact with a computer (Chapter 2). A great disadvantage, however, is random variability at the expense of experimental control. Even when the confederate's speech is scripted (e.g., Branigan et al., 2000), its acoustic properties, such as the strength of their accents and the degree of accommodation to the participants, are completely unpredictable (Lewandowski & Nygaard, 2018).

Conversely, computer confederates have the advantage that they enable to study interactions in a fine-grain controlled setting with the use of pre-recorded speech (e.g., Cohn et al., 2019; Cohn & Zellou, 2019; Eijk et al., 2019; Gessinger et al., 2019; Raveh, Siegert, et al., 2019; Raveh, Steiner, et al., 2019). Thus, researchers can assess the effect of various factors on phonetic accommodation in an interaction, for instance the expressiveness of the voice of the confederate (Cohn & Zellou, 2019), the frequency and neighborhood density of the words included in the interaction (Pardo et al., 2013), the type of feedback that participants receive from the interlocutor and even the exposure to a specific variant of a sound included in the speech of the confederate (Gessinger et al., 2019). Among the disadvantages of using computer-based approaches is that the implementation of computer-based approaches often leads to the simplification of the task, sacrificing the interactiveness and spontaneity of the conversation. For instance, word reading (Babel et al., 2013; Pardo et al., 2012) and shadowing tasks (Cohn et al., 2019; Lewandowski & Nygaard, 2018; Pardo et al., 2013; Rojczyk, 2013) are frequently used with computer-based approaches. The oversimplification of an interaction yielded can have consequences on the conclusions of the studies investigating interactions. For instance, when studying conversation-based phenomena such as phonetic alignment in a laboratory setting, the use of tasks lacking the conversational goal of an interaction (Pardo et al., 2017) could be counterproductive. Similarly, the use of shadowing tasks, a task which could risk inducing plain imitation (Pardo, 2006), could also lead to conclusions that are not representative of daily life interactions. In fact, the speech samples recorded for this type of settings are conversationally poor, sometimes limited to single-word utterances (Cohn et al., 2019; Gessinger et al., 2019; Liu & Johnson, 2017; Pardo et al., 2013). Furthermore, when using computer-based approaches, the humanness of the confederate is often reduced to a human voice (Babel, 2012; Cohn et al., 2019) or a human-looking virtual avatar (Staum Casasanto et al., 2010) but participants are not supposed to believe they are talking to a real human being.

Given the experimental control allowed in computer-based approaches, some researchers have used this type of method to study human-human interactions, based on the implicit assumption that the results from computer-based approaches are also representative of human-human interactions (e.g., Babel, 2010, 2012; Cohn et al., 2019; Eijk et al., 2019). However, others have emphasized that these two settings should be treated differently (Moore, 2016), because the literature comparing human-computer and human-human interactions has provided overwhelming evidence showing significant differences between interactions involving only humans and interactions where humans talk to a computer (Branigan et al., 2010, 2003; Burnham et al., 2010; Cohn et al., 2019; Gessinger et al., 2017; Raveh, Steiner, et al., 2019).

One problem with the comparison of human-human and human-computer interactions is that these studies take place in a non-spontaneous context, which hinders the generalizability of the conclusions (Branigan et al., 2003, 2010; Burnham et al., 2010; Cohn et al., 2019; Gessinger et al., 2017; Raveh et al., 2019). One exception is Burnham et al. (2010), where the two setting are compared in using an interactive task. A limitation in that study is that the phonetic control over the input varied across the two settings, potentially affecting the comparison. The current chapter presents a direct comparison between human-human and human-computer interactions where the speech input is kept constant between the two settings, and the interaction is based on a task preserving the conversational goals and spontaneity. In addition, in the human-human setting, participants are expected to be believe that they are having a conversation with a real human being.

3.1.2. The challenge of comparing human-computer vs human-human interactions

Among the factors that modulate the presence of phonetic accommodation, researchers have reported social aspects, such as the prestige attached to the variety spoken (Willemyns et al., 1997), the role that the speaker performs during the conversation, for instance giving or receiving instructions (Pardo, 2006), the gender of the interlocutors (Cohn et al., 2019; Namy et al., 2002; Pardo, 2006), the likeability (Babel, 2012) and the hierarchy status (Willemyns et al., 1997). Factors that are intrinsic to the language have also been found to be relevant to the degree to which speakers adjust their pronunciation in an interaction (Berry & Ernestus, 2018; Delvaux & Soquet, 2007; Goldinger, 1998; Hwang et al., 2015; MacLeod, 2014; Pardo et al., 2013). For instance, MacLeod (2014) observed that, when two interlocutors differ in the pronunciation of a given phonetic cue, the more perceptually acoustically distant the difference is, the more likely speakers will align to the interlocutor's pronunciation of that given cue.

Although these same factors may be relevant in the two settings, their weight may differ depending on whether or not the interlocutor is human (Branigan et al., 2010; Burnham et al., 2010; Cohn et al., 2019;

Raveh, Siegert, et al., 2019; Staum Casasanto et al., 2010). In fact, theoretical approaches on computer personification, for instance in the Computer As Social Actors framework (Nass & Moon, 2000), predict that a human speaker may show different behavior when presented with a computer interlocutor depending on the degree of humanness this computer shows (e.g., voice versus voice and avatar). According to these predictions, speakers would show phonetic accommodation patterns more similar to those found in human-human interaction, the more human-like the computer is presented. Phonetic accommodation in human-human settings seems to be more affected by mechanisms such as priming, which are intrinsic to the conversation rather than mediated by the beliefs the speaker has about the interlocutor, than phonetic accommodation in human-computer settings (Branigan et al., 2010). For instance, for speakers interacting with another human, priming resulting from mere exposure has been found to trigger more phonetic accommodation, in particular phonetic alignment, from human voices than from computer voices (Branigan et al., 2010; Cohn et al., 2019).

In conversations involving a human and a computer, on the other hand, communicative success seems to have a primary role in the conversation. Speakers tend to adapt their pronunciation in conversation with computers in order to ensure the understanding from the interlocutor, which has limited abilities to accommodate to the speaker (Branigan et al., 2010; Burnham et al., 2010). Following up on this, the current study focuses on how mere exposure and communicative success affect phonetic accommodation to humans and to computers differently by directly comparing the phonetic accommodation present in human-human and human-computer settings when speakers receive mere exposure to the interlocutor's pronunciation of given segments and feedback implying that the communication is not successful.

3.1.2.1. The effect of exposure onphonetic accommodation

There is general consensus that exposure to the speech of the interlocutor is essential for phonetic accommodation to take place. Some researchers have even argued that priming alone is the driving force of phonetic accommodation (Pickering & Garrod, 2004). For instance, in terms of phonetic alignment, when speakers hear the interlocutor produce a cue, they align to that cue because the exposure to that cue boosts the activation level of that variant (Pickering & Garrod, 2004). In order to study the effect of priming on phonetic accommodation, research has focused on how the presence of exposure only can trigger speakers adapting their pronunciation of not only in the L1 (Babel et al., 2013; Delvaux & Soquet, 2007; Goldinger, 1998; Namy et al., 2002; Pardo et al., 2012) but also in the L2 (Berry & Ernestus, 2018; Hwang et al., 2015). In human-human interactions, Pardo et al. (2006) found that speakers tended to align the pronunciation to the interlocutor after briefly being exposed to the interlocutor's speech in a word-reading task. In the context of L2, Hwang et al. (2015) found that Korean speakers of English adapted their pronunciation of problematic English contrast to sound more native-like after interacting with a native speaker of English. Similarly, in computer settings, studies using computer-based approaches where participants had to interact with a computer have also found an effect of exposure on phonetic accommodation (e.g., Babel et al., 2010, 2013).

Although the exposure can trigger phonetic accommodation in both human-human and human-computer interactions, the weight of this factor differs depending on how human-like the interlocutor is (for a review see Branigan, 2010). There are clear tendencies for exposure to trigger more phonetic accommodation in human-human interactions. For instance, Cohn et. al. (2019) found that participants, who had to shadow human and computer voices to which they were exposed, showed more phonetic accommodation matching the human voice even though they received the same amount of exposure to the voice of each talker. These results confirm that phonetic accommodation cannot only be the result of mere priming but suggests that phonetic accommodation has to be a socially-mediated process (Cohn et al., 2019). The current study explores the effect of exposure on phonetic accommodation in human-human and human-computer interactions.

3.1.2.2. The effect of feedback on phonetic accommodation

It is commonly argued nowadays that phonetic accommodation is the result of a speaker's efforts to ensure that the interlocutor will be able to understand the message transmitted (Giles & Ogay, 2007). The degree of phonetic accommodation is positively correlated with the degree of success in the performance (Reitter & Moore, 2014).

In conversations with non-native speakers the risk of misunderstanding is large because the distance between the pronunciation of the interlocutors is maximal (Cenoz & García-Lecumberri, 1999). The mismatches between the sound inventories of different languages can lead to substantial differences between the native and the nonnative pronunciation of a given sound. As a consequence, non-native speakers could align to a native interlocutor's pronunciation in order to avoid disruptions in the conversation triggered by these mismatches. Feedback from the interlocutor is a clear way to alert non-native speakers of the distance between their production and the native-like form of a given sound (Mackey & Gass, 2006).

Feedback has been found to be overall effective in triggering adaptations in speakers' production (Aliaga-García & Mora, 2008; Burnham et al., 2010; Dlaska & Krekeler, 2013; Kartushina et al., 2015; Saito & Lyster, 2012). However, its effectiveness is closely linked to the nature of the feedback (Adams et al., 2011; Mackey & Gass, 2006). The feedback speakers can receive on their pronunciation can vary in different dimensions. First, the explicitness of the feedback can be manipulated to be overt and explicit (Kartushina et al., 2015; Thorin et al., 2018) to more subtle and implicit feedback (Burnham et al., 2010; Stent et al., 2008). In the context of phonetic adaptations during interactions, research focuses on this last type of feedback, which seems to be more suitable during conversation (Burnham et al., 2010). In addition, feedback can also vary in whether it is presented in isolation or with positive evidence of the model pronunciation, i.e., examples of the pronunciation of the interlocutor (e.g., Adams et al., 2011; Dlaska & Krekeler, 2013; Mackey & Gass, 2006; Saito & Lyster, 2012).

Interestingly, whereas the effect of exposure on phonetic accommodation has been found to be larger in human-human than in human-computer settings, the opposite has been found for feedback. Burnham et al. (2010) presented feedback using a scripted confederate, either with a computer interlocutor with a facial avatar, or a human interlocutor presented in a video. The feedback consisted of the confederate pretending to misunderstand the corner vowels (/i, a, u/) produced by the participants and it accounted for 25% of the sentences. The feedback trigged more adaptation when speakers interacted with the computer interlocutor than with the human confederate (Burnham et al., 2010). The present study aims to expand the comparison of the effect of feedback on phonetic accommodation in human-human and human-computer settings.

3.1.2.3. The combination of feedback with exposure and phonetic accommodation

Studying how feedback may differ if presented in combination with exposure when speakers interact with a human confederate requires more experimental control than possible in regular human-human conversations. Studies in the field of applied linguistics have aimed to investigate the effect of these two factors separately by comparing the effect of feedback with and without exposure to the word of interest. For instance, in the L2 classroom, corrective feedback can be presented as "recasts", where the interlocutor repeats the word providing the speaker with exposure to the nativelike pronunciation of the sound, or as "prompts", where the interlocutor asks for a repetition of the word implying that the learner's pronunciation was unclear, without exposing the speaker to the native-like pronunciation of the sound. (e.g., Gooch et al., 2016; Saito & Lyster, 2012). For example, Gooch al. (2016) shows that Korean learners of English produced more target-like pronunciations of the English /ɹ/ after re-casts than after prompts. However, it is impossible to control exposure to specific sounds with a human interlocutor.

Given that exposure and feedback in isolation affect phonetic adaptation differently in human-human and human-computer settings, it is conceivable that exposure and feedback when presented together may also interact differently with each other in human-human and human-computer. The current study aims to investigate whether the combination of exposure and feedback affects phonetic adaptation differently in human-human and human-computer contexts.

3.1.2.4. The Ventriloquist Paradigm

Recent developments in computer-based approaches have allowed researchers studying the role of exposure and feedback to move away from shadowing tasks and implement more interactive methodologies in their experiments. One well-attested paradigm used to achieve this aim is the Wizard of Oz, where participants interact with a smart computer, which is in reality controlled by a researcher playing pre-recorded sentences to the participant (Riek, 2012). For instance, in Gessinger et al. (2019), German speakers interacted with a computer interlocutor, called Mirabella, who could ask and answer questions, in the style of a virtual assistant. Thus, more interactive tasks can be implemented in controlled settings. These tasks normally involve question-answer games that attempt to preserve the conversational goal, for instance with prescribed question-answer games (Burnham et al., 2010). Besides the Wizard of Oz other methodologies have also attempted to implement more human-like confederates, for instance by making use of virtual reality avatars (Staum Casasanto et al., 2010) or screen-presented scripted confederates (Burnham et al., 2010). However, in order to achieve a completely balanced comparison between human-human and human-computer interactions, researchers would need a paradigm that allows for the combination of tasks preserving the conversational goals allowing for spontaneity, the presence a human confederate in the same room where the participant is, and the implementation of the exact same speech materials in the two settings to ensure participants are exposed to the same input in both settings.

Recently, we have developed a new experimental paradigm which aims to combine experimental control and ecological validity of interactions: the Ventriloquist Paradigm (Chapter 2). This paradigm is the first method that allows for the covert implementation of prerecorded speech in an interaction where participants think they are having a conversation with a human confederate sitting across the table from them. In reality, the confederate is playing prerecorded speech samples unbeknownst to the participants. With this methodology, the participant and the human confederate, who are sitting in the same room, are equipped with headphones and microphones to interact with each other, for instance while playing a cooperative puzzle-solving task together. The participant is told they will be hearing each other through the headphones but everything the participant hears is prerecorded speech. During the interaction, the confederate uses a hidden numeric keypad to play speech samples and subtly hides her face behind a monitor every time a speech sample is played.

In addition, the second innovative aspect of this paradigm is successful implementation of the pre-recorded speech while preserving the spontaneity of the conversation. The pre-recorded speech samples include spontaneous categories, for instance, affirmative and negative responses, requests for more information from the participant, reassuring utterances and backchannel sounds, among others (see Chapter 2 for a detailed overview). The implementation of a set of categories of "spontaneous" speech utterances, which can be played during the whole experimental session, helps maintaining the illusion of spontaneity in the interaction. This aspect is essential in order to preserve the ecological validity in the interaction. The combination of the possibility of implementing pre-recorded speech and the preservation of ecological validity of the interaction makes the Ventriloquist Paradigm a most suitable methodology to compare human-human and human-computer settings in interaction. In the present study, we use this paradigm to implement the same pre-recorded speech to study phonetic adaptation in human-human and human-computer settings, in a highlyinteractive context.

3.1.3. This study

This study investigates whether the degree of phonetic adaptation that speakers show in an interactive setting depends on whether the interlocutor is presented as a human confederate or a smart computer, even when the speech of the interlocutors is identical in both settings. In order to do so, native speakers of Dutch were asked to play a puzzle-solving game in English with an English native speaking interlocutor. Half of the participants believed they were interacting with a human being sitting across them in the same room and the other half believed they were interacting with a computer. We analyzed their production of the English vowel contrast in "bad-bed" before and after the interaction. This contrast, which in English is distinguished in terms of height and frontness, has been reported to be difficult for this population of L2 speakers (Elsendoorn, 1985; Wang & van Heuven, 2006), as they tend to merge these two vowels in an in-between sound. Therefore, phonetic adaptation was measured as the degree to which participants un-merged this English contrast in the native-like way.

The novelty of this study lays in two aspects. First, unlike previous research that has contrasted these two settings in contexts where the conversational goal was removed from the interaction (e.g., Cohn et al., 2019; Gessinger et al., 2017; Liu & Johnson, 2017; Pardo et al., 2017), in the present study the comparison is based on a spontaneous cooperative interaction between the participants and the interlocutors. Second, the comparison between the two types of interactions, human-human and human-computer, is established while maintaining the two settings identical: in both settings the same materials, including all auditory input participants hear, are used. In the human-human setting, the interlocutor was presented as a human confederate sitting in the same room as the participant. However, the human confederate was actually playing pre-recorded speech, unbeknownst to participants, using the Ventriloquist Paradigm (Chapter 2). In the computer-setting, participants believed they were interacting with a computer in the style of the Wizard of Oz (Riek, 2012). In reality, the computer's contributions to the interaction were controlled by the same human confederate as in the human-human setting, using the exact same pre-recorded speech. Using pre-recorded speech allowed for the full control over the exposure participants received to the interlocutor's pronunciation of the specific sounds during the interaction.

This study investigates the differences between phonetic adaptation in human-human and human-computer interactions. In particular, the focus of the study is to determine whether the presence of exposure and feedback, in isolation and in combination, affects phonetic adaptation differently in human-human than in human-computer settings, when everything else including the speech of the interlocutor is kept identical across the settings. Thus, our research questions are:

> RQ1: Does exposure to the interlocutor's pronunciation of the sounds of interest trigger phonetic adaptation differently in human-human interactions than in human-computer interactions?
>
> RQ2: Does feedback on the speakers' pronunciation of the sounds of interest trigger phonetic adaptation differently in human-human interactions than in human-computer interactions?
>
> RQ3: Does the combination of feedback on the speaker's pronunciation of the sounds of interest and exposure to the interlocutor's pronunciation of the same sounds affect phonetic adaptation differently in human-human interactions than in human-computer interactions?

In line with previous work using less interactive tasks (Burnham et al., 2010; Cohn et al., 2019; Raveh, et al., 2019), we expect overall differences in phonetic adaptation in the human-human and human-computer settings. In order to assess the effects of exposure to and feedback on the sounds of interest on phonetic adaptation during the interaction, a baseline condition without such exposure and feedback was also incorporated; any differences found in this condition will result from merely participating in the interaction.

In general, given that phonetic adaptation tends to be a socially-mediated process, participants in the human-human setting could be overall expected to show a higher degree of phonetic adaptation than those who interact with a computer. Following previous research which argues that phonetic adaptation is more likely to be affected by exposure in human-human settings (Branigan et al., 2010; Cohn et al., 2019), participants who interact with the human confederate are expected to show a higher degree of adaptation when exposure is present during the interaction than those who receive the exposure from the computer (RQ1). As regards with feedback, since communicative success tends to be more relevant in human-computer phonetic adaptation than in human-human settings (Branigan et al., 2010), feedback implying misunderstanding could trigger more adaptations when speakers interact with a computer than when they interact with a human interlocutor (RQ2). Similarly, the effect of combination of the two factors together, exposure and feedback, on phonetic adaptation in these two settings could also be expected to trigger greater phonetic adaptation from participants when they interact with a computer than with a human confederate, following previous studies (RQ3) (e.g., Burnham et al., 2010).

The current study focuses on analyzing the effects of exposure and feedback on the participants' long-term pronunciation. Therefore, the scope of the study involves the measurement of the general adaptation of the sounds of interest, by comparing the pronunciation before and after the interaction, rather than local adaptation in the participants' pronunciation immediately after receiving exposure or feedback.

The contrast investigated in this study is the English vowel contrast, "bad"-"bed", which Dutch speakers tend to merge into an in between sound closer to the vowel in English "bed" than in "bad". Following the accounts on how the perceptual salience of a difference affects phonetic adaptation (MacLeod, 2014), we expect the vowel that is more different from the native pronunciation, "bad" to undergo more adaptation than the nonnative vowel in "bed", which is closer to the native pronunciation. These two vowels differ in terms of height and frontness in English, however, Dutch speakers are more proficient at distinguishing these two vowels in the height dimension than in the frontness dimension (Troncoso-Ruiz et al., 2019). Due to this asymmetry, the two dimensions might be affected differently when speakers try to adapt their pronunciation. For instance, participants may find it easier to enlarge the height distinction that they already used for distinguishing the vowels to start with, than to learn to distinguish these two vowels in the 'new' frontness dimension.

3.2. Methods

In order to answer the research questions, we designed an experiment where participants interacted with either a computer or a human confederate while playing a cooperative puzzle-solving game. We used the Ventriloquist Paradigm in the human-human setting in order to covertly incorporate prerecorded speech during the interaction and, a computer-implementation of the same paradigm for the human-computer setting. This interaction was preceded and followed by identical production reading tasks as pre-test and post-test. A dummy tone discrimination task was used to avoid any contact between the participants and the human confederate before the interaction in the human-human setting, and in the human-computer setting for comparability.

Between participants, exposure (presence/absence), feedback (presence/absence), and setting (human-human/human-computer) where fully crossed, resulting in eight groups. Exposure to the vowel contrast under study was fully controlled using prerecorded speech. Feedback was implemented as implicit visual negative feedback on the pronunciation of this contrast by the participants.

The following sections will discuss the demographic characteristics of the participants (section 3.2.1), the materials used in the experiment (section 3.2.2.) and the procedure of the experiment (section 3.3).

3.2.1. Participants

Ninety-six native speakers of Dutch participated in this study (age: M= 21.5 years, SD= 2.28 years). All were female, as gender has been shown to affect the degree of phonetic accommodation (Cohn et al., 2019; Namy et al., 2002; Pardo, 2006), and including gender as a factor was outside the scope of this study. All studied at Radboud University Nijmegen or HAN University of Applied Sciences.

We recruited students with an intermediate level of proficiency of English to ensure they would have room for improvement in their pronunciation. Since Dutch speakers are usually relatively fluent in English and receive an average of 6 to 8 years of English instruction in primary and secondary education, we recruited participants who did not have daily contact with English. Therefore, only students who were not enrolled in English-taught programs were invited to participate. Participants' average score on the LexTale test (Lemhöfer & Broersma, 2012) was 75.84 (sd = 10.16), which corresponds to an upper intermediate (B2) level in the Common European Framework of Reference for Languages (Council of Europe, 2001). None of them reported any hearing, reading or learning impairments and they were compensated for their participation in the study with a voucher or study credits. Forty-eight participants were assigned to the humanhuman setting, where they interacted with a human confederate, and the other 48 were assigned to the human-computer setting, where they interacted with a computer.

3.2.2. Materials

3.2.2.1. Minimal pairs

Five English sound contrasts were selected, one experimental contrast that was very difficult to produce for this population, namely /æ, ɛ/, and four filler contrasts that were relatively easy for this population, namely /ɪ, i:/, /p, b/, /m, n/, and /r, l/. Filler contrasts were included to divert attention from the experimental contrast. Both the critical contrast and the filler contrasts were always presented in the context of minimal pairs (see section 3.2.2.4.). For each of the five contrast, 24 minimal pairs were selected consisting of high-frequency words that the participants were expected to know. The minimal pair words containing the contrast /æ-ɛ/ were considered the critical. The average *zipf* value of the words including /æ-ɛ/ was 4.81 (SD =.661) in a 1-7 scale, according to SUBTLEX-UK (van Heuven et al., 2014). All of the items were monosyllabic except for "cattle"-"kettle" and "marry"-"merry"(see Appendix 1 for the full list).

3.2.2.2. Sentence completion task

For each of the 24 critical minimal pair words, for instance "bad", two matched sentences were created including the critical word (e.g., (1) "You have some **bad** habits, such as…" and (2) "I have some **bad** news, you…"). The utterances varied in length and difficulty to complete, but the sentences including the same critical word were designed to be as similar as possible. These keywords were included to make sure participants produced these sounds before and after the interaction in similar contexts. Thus, a total of 96 incomplete utterances were generated to be used as material for the sentence completion tasks. Twelve different lists were created in order to counterbalance the order in which the matched sentences containing the same critical word appeared in the pre-test and post-test tasks.

3.2.2.3. Distraction Task

Thirty pure tones were created using Praat (Boersma & Weenink, 2019) for a tone discrimination task played between the pre-test sentence completion task and immediately before the interaction. The task consisted of 15 trials which contained two tones presented together. The frequency of tones varied from 400 to 765Hz. The difference between the tones included in each trial ranged from 0 to 295Hz. A single randomized list was used for all participants.

3.2.2.4. Cooperation Game/Interaction Stage

The interaction between the participants and the confederate was based on a cooperative game of 77 trials belonging to two different categories: 64 Code Breaker trials, designed to manipulate the presence/absence of feedback, and 13 Semantic Relation trials, designed to manipulate the presence/absence of exposure to the experimental contrast. The game did not have a time limitation, and it was paced according to the performance of the participant and the interlocutor.

A total of 24 lists were used for this part of the experiment. First, twelve different lists were created using the same structure (see Figure 1). Each list started with the same block of practice trials including four Code Breaker trials and one Semantic trial, followed by 12 blocks of test trials, each block containing five Code Breaker trials and a Semantic Relationship trial. One out of the five Code Breaker trials was a critical trial. Each list varied in the order in which the critical code breaker trial was presented in each block (i.e., first, second, third, fourth or fifth position). All the Semantic Relationship trials were considered critical and all of the trials included the critical contrast. These 12 lists were used in the conditions with the presence of exposure. Twelve additional lists were created as duplicates of the original 12, where the Semantic Relationship trials did not include the critical contrast. These last 12 lists were used in the conditions without exposure.

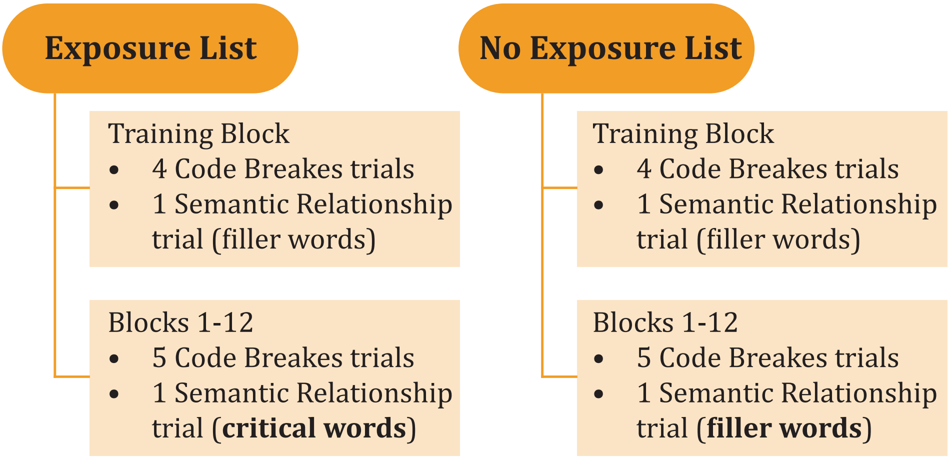


Figure 1 Visual representation of the overall structure of the lists used for this study

Feedback manipulation: Code Breaker trials

For each Code Breaker trial, two screens were designed (see Figure 2). The top half of the confederate screen showed a sequence of geometrical shapes designed to follow an identifiable pattern. Below the sequence, a set of two minimal word pairs was displayed (Calibri, 28pts). The words on the screen belonged to the battery of minimal pairs selected for this study (see section 3.2.2.1). The participant screen contained four different individual shapes, among which only one matched the pattern in the sequence on the confederate screen. The same subset of four words from the confederate screen appeared in the participant screen, each word appeared below one of the shapes. The minimal pair which included the word below the right shape was considered the target minimal pair and the other one was the background minimal pair.

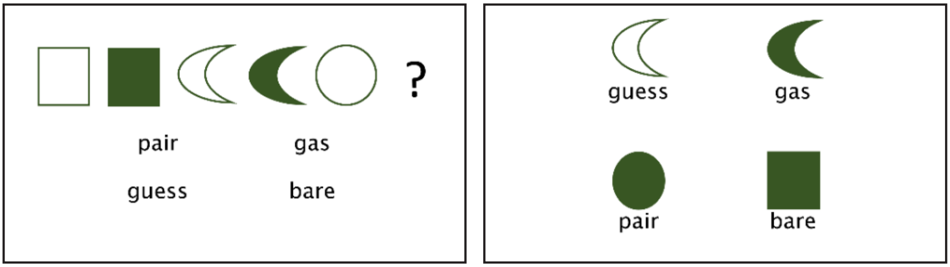


Figure 2 Confederate screen (left) and participant screen (right) as displayed during the Code Breaker trials.

Each of the 24 lists contained: 4 training trials, 48 filler trials and 12 critical trials of the Code Breaker type. In the critical trials, the target minimal pair represented the /æ, ɛ/ contrast and the word linked to the right shape always contained the /æ/ vowel. The location in which the critical minimal pair words appeared on the screen was randomized, to avoid the critical contrast to appear in the same position in all critical trials. In each trial, the position of the same words in the two screens was different (see Figure 2). In the trials that were not considered critical trials, the filler trials, the target minimal pair included one of the other four filler contrasts. Thus, in these trials, the word linked to the right shape never included the critical vowels, but filler words. In these filler trials, the /æ, ɛ/ contrast could appear as the background minimal pair words.

Exposure manipulation: Semantic Relation trials

Each of the 24 lists also contained 13 Semantic trials: 1 training and 12 critical trials. The 13 Semantic Relation trials also consisted of two screens (see Figure 3). The confederate screen showed two minimal pairs, whereas the participant screen displayed a picture. The picture in the participant screen was related to one of the words displayed in the confederate screen.

Figure 3. Confederate screen (left) and participant screen (right) as displayed during the Semantic Relation trials.

Twenty-four lists were generated for these trials randomizing the order of appearance of the pictures. Out of the 24 lists, 12 included /æ, ɛ/ minimal pairs among the words displayed on the participant screen. These 12 lists were used in the Exposure Present conditions. The critical minimal pairs included in these lists did not appear in the Code Breaker trials, and the order of the words (i.e., the word containing /æ/ or /ɛ/) was counterbalanced and randomized. The critical minimal pair were never the words related to the picture on the participant screen. The other 12 lists were identical to the lists used in the Exposure Present lists, with the only difference being that the four words included in the participant screen belonged to the filler minimal pairs selection.

3.2.2.5. Recordings for the ventriloquist voice

A 25-year-old female speaker of British English was recorded as the voice of the confederate in a recording booth, using a Senheisser K3 microphone and a Roland R09HR recorder. A female speaker was chosen in order to match the gender of the participants and avoid gender effects on phonetic accommodation. She was recorded reading a diverse battery of sentences, which were carefully scripted avoiding the /æ, ɛ/ vowels. The battery of sentences consisted of two main categories: (1) 257 sentences linked to the information displayed on the screen in every trial during the interaction, e.g., "So, all of the shapes are smiley faces", and (2) 86 spontaneous speech utterances. This second set of recordings, the spontaneous speech utterances, was designed to help the confederate answer questions and react to the participants' spontaneous behavior, for instance, when participants asked questions or requested more information or clarification. The spontaneous categories included were: affirmative responses, negative responses, backchannel sounds, variations of "I don't know", requests for clarification, reassuring utterances, reminders about the rules of the task and utterances showing surprise for a new trial starting. In order to make these utterances believable and natural, different versions of the same utterance, for instance "no", were recorded with differences in intonation or duration (see Chapter 2 for a complete description of the spontaneous speech categories used).

The 257 sentences linked to the information on the screen could be related to either the Code Breaker trials or the Semantic Relation trials. For each Code Breaker trial, three audios were recorded: one included a description of the sequence of shapes displayed on the confederate screen (e.g., "So all of the shapes are smiley faces, I think the word we're looking for should be linked to a pink smiley face"). The other two additional audios contained similar information phrased differently and they could be played in case the participants asked for a repetition (e.g.,"I think we need the word linked to a pink smiley face and do you see a word linked to a pink smiley face in your screen?").

In the audios for the Semantic Relation trials, the speaker read the words displayed on the screen and chose the one that was most closely related to the picture displayed on the other screen (e.g., "the words on my screen are bad, bed, motion and notion, so I think it is motion”). Filler trials only had one audio attached to them while two versions were recorded for critical trials containing the /æ, ɛ/. The versions differed in the order in which the contrast was presented (/æ, ɛ/ vs /ɛ, æ/).

3.3. Procedure

For this study, a between-participants design was selected. The total of 96 participants recruited for this experiment were divided into two groups: 48 participants interacted with the human interlocutor and the other 48 interacted with the computer interlocutor. In each of the two settings, the participants were randomly assigned to one of the four conditions of the experiment: Control, OnlyExposure, OnlyFeedback and ExposureAndFeedback. The procedure in both settings was identical: participants first performed a pre-test, followed by the distraction task (meant to prevent them from talking to the confederate in the human-human setting), then the interaction, and finished the experiment with a post-test and questionnaires.

3.3.1. Pre-test

3.3.1.1. Sentence Completion Task

The experiment session started with participants performing a Sentence Completion Task, used as the pretest. At the beginning of the experiment, before the pretest Sentence Completion Task, participants were told that, due to its length, the task would be split into two parts. Thus, the pretest-posttest structure was somewhat disguised for participants. The sentences were presented as a PowerPoint slideshow. Only one incomplete sentence was presented in each slide (Calibri, 28pt). The experiment was self-paced, so participants could choose a comfortable speed to perform the task. They were instructed to, first, say the ending that could complete the sentence and then read the whole sentence immediately followed by the suitable ending out loud to ensure fluency when reading the sentence.

3.3.1.2. Distraction task

The aim of this task was to create a distraction for the participants while the human interlocutor, the ventriloquist, entered the room where the experiment took place. This distraction prevented participants from trying to speak to the confederate before the confederate was equipped with the headphones and playable pre-recorded speech. Participants in the human-computer setting also performed this task in order to keep the two experimental sessions as identical as possible. Participants were equipped with headphones and a button box to perform this task, implemented on Matlab. They were exposed to two tones through the headphones and were asked to indicate whether the tones were different using a button box.

3.3.2. Interaction

3.3.2.1. Settings

Human-Human Setting

In the human-human setting, participants sat in a big recording studio across the table from the human confederate. Prior to the interactions, participants were told they would be playing a game in English with a native speaker. No motivation was given as to why an English-speaking person was involved. Both the participant and the confederate were equipped with headphones and microphones. In this setting, the pre-recorded voice was embodied by a human confederate. Participants were led to believe the audio they could hear through the headphones was coming live from the human confederate microphone while, actually, the audio had been pre-recorded. The confederate used a hidden numeric keyboard to play pre-recorded clips to the headphones of the participant and used the monitor on the table to hide her mouth.

Human-Computer Setting

In the human-computer setting, the interaction took place in a smaller recording booth where participants were also equipped with headphones and a microphone and performed the task on a real computer. Before the interaction, participants were told they would play a game with a smart computer and, identically to the human-human setting, no motivation was given as to why the computer spoke English. In this setting, a variant of the Ventriloquist Paradigm was used in order to make participants believe the interlocutor was a smart computer similar to a virtual assistant (like Siri or Alexa) in the style of the Wizard of Oz (Riek, 2012). This version of the paradigm is identical to the original Ventriloquist Paradigm, the only difference is that the pre-recorded voice is embodied by a real computer located in the same booth as the participants. In this version of the experiment, the same confederate as in the human-human setting was responsible for playing the pre-recorded speech, using the same range of pre-recorded utterances. However, in the human-computer version, the interlocutor presented in the same room as the participants was the computer and the confederate sat in a different room out of the sight of the participant.

3.3.2.2. Code Breaker

In the Code Breaker game, the participant and the confederate, human or computer, cooperated sharing information on their screen with each other in order to solve puzzles presented on the screens. Every Code Breaker trial started with the confederate playing an audio describing a sequence of shapes. The participant then tried to find out which of the shapes on their screen could match the sequence described by the interlocutor. Once both the participant and the confederate agreed on which shape was the right one, the participant instructed the confederate to click on the word below that shape.

In the Feedback Absent conditions, the trial ended when the confederate clicked on the word pronounced by the participant and the next trial appeared on the screen. In the Feedback Present conditions, after the confederate clicked on the word, participants could see a gray rectangle appearing around the word selected by the confederate for 3.0 seconds. In filler trials, the confederate chose the word the participant instructed her to click on. In critical trials, where participants had to instruct the confederate to click on a word containing /æ/, for instance “bad”, the confederate always clicked on the other member of the minimal pair containing the /ɛ/ vowel, “bed”.

The feedback remained corrective throughout the whole experimental session regardless of the actual pronunciation of the participants in order to maintain the experience of all participants as homogenous as possible. Note that the proportion between critical trials where participants received corrective feedback and filler trials was kept low: only 20% of the experimental Code Breaker trials were critical.

3.3.2.3. Semantic Relation Trials

In the Semantic Relation trials, participants started describing the picture displayed on their screen. The confederate played an audio file telling the participants which words were displayed on her screen and choosing the one that seemed most directly related to the picture described by the participant. The trial ended with the confederate clicking on that word. Before playing the audio reading the answer, the confederate could also play audio files asking the participant to elaborate on the description.

3.3.3. Post-test

3.3.3.1. Sentence Completion Task

After the last Semantic Relation trial, participants received a notification on the screen asking them to remain seated while the confederate played an audio announcing she had to leave the room to perform another task. The participant then received another notification on their screen asking them to finish the second part of the Sentence Completion Task they had started at the beginning of the experimental session. Since participants were already familiar with the instructions for this task, they could perform this task without interacting with the experiment leader.

3.3.3.2. Language Background Assessment and interaction-related questionnaires

After participants finished the post-test Sentence Completion task, they did the English online version of the Lexical Test for Advance Learners of English (LexTale)(Lemhöfer & Broersma, 2012). This visual lexical decision task intends to provide researchers with an objective measurement of participants' proficiency level. Next, participants filled in a questionnaire which included questions about their language background, age, gender and the study program they followed. In this questionnaire, participants were also asked to rate the friendliness of the interaction and how well the other player, human or computer, could understand them on a five-point Likert Scale. Finally, participants received an oral debriefing from the experimenter leading the experimental session. This debriefing included questions which were meant to elicit whether participants had noticed the pre-recordedness of the speech of the confederate. This debriefing also included information about the feedback being pre-programmed.

3.4. Results and discussion

3.4.1. Outcomes of the questionnaires

In order to determine whether the confederate playing the pre-recorded utterances behaved differently in the human-human and the human-computer settings, the number of played utterances were compared across settings using the “lm” function of the *lmerTest* package (version 3.1.0) (Kuznetsova et al., 2017) in R (version 3.5.0) (R Development Core Team, 2008). The number of utterances used in each of the settings was not found to be significantly different between the human-human setting (M = 85.40, sd =4.19) and the human-computer setting (M = 85.96, SD = 4.47) (β=-0.56, t=-0.665). This finding confirms that the confederate behaved similarly in terms of the pre-recorded speech utterances in the two settings.

In order to investigate whether participants in the human-human setting believed they were having a live conversation with the human confederate, their responses to the oral debriefing about the pre-recordedness were analyzed. Out of the total 48 participants who interacted with the human confederate, 75.0% reported no suspicion about the speech of the interlocutor being pre-recorded. This outcome indicates that the use of the pre-recorded samples was used convincingly in most of the interactions. Participants who did report noticing the pre-recordedness were included in the data set because the analyses did not show differences between participants who believed the interlocutor was speaking live and those who did not.

In the last stage of the experiment, participants used a five-point Likert scale to rate the friendliness of the interlocutor, computer or human. The mean rating for friendliness was relatively high in both settings: in the human-human setting, the mean value was 4.54 (sd = 0.582); and in the human-computer setting the mean rating was 4.53 (sd = 0.581). In order to investigate whether the setting and the condition to which participants were assigned could predict their ratings on the scale, their responses were analyzed using ordinal regression with the “ordinal” package (version 2012.12.10) (Christensen, 2019) in R (version 3.5.0) (R Development Core Team, 2008). The model included the rating as the outcome variable and Setting (human-human/Human-computer), Condition (Control, OnlyExposure, OnlyFeedback, ExposureAndFeedback) and their interaction as the fixed factors. The model outcome (see Appendix 2) showed no significant differences between the conditions or the settings.

Similarly, participants also rated how well the other player, human or computer, understood them during the interaction. The mean rating for the intelligibility of the human interlocutor was 3.77 (sd = 0.555), whereas the mean rating for the intelligibility of the computer interlocutor was 3.90 (sd = 0.653). Again, ordinal regression was used in order to investigate differences among the two settings and the conditions, with the rating as the outcome variable and Setting, Condition and their interaction as the fixed factors. The results (see

Appendix 3) also showed no differences among the conditions or the settings.

3.4.2. Data Preparation

The 96 sentences that participants produced in the Sentence Completion Tasks were automatically forced aligned using the Montreal Force Aligner (MFA) (McAuliffe et al., 2017). This software aligns audio input to its orthographic transcription using acoustic phone models and a dictionary mapping words to their pronunciation specified as a sequence of phones. Since there are no models for Dutch-accented English available, the default MFA pre-trained phone models for English based on the LibriSpeech dataset (Panayotov et al., 2015) were used for the alignment of these sentences. The Carnegie Mellon University dictionary (Weide, 1994) was updated so as to include all the words in the utterances recorded. Furthermore, the phone transcriptions available in the dictionary were expanded to include possible Dutch-accented English pronunciation in order to improve the alignment. For instance, the dictionary included two transcriptions for the word "bid": the native-like English form, [bɪd], and the Dutch-accented version including final devoicing, [bɪt].

The first (F1) and second (F2) formant values were extracted from the midpoint of every token of /æ/ and /ɛ/ that participants produced in the pre-test and in post-test sentences. The tokens with formant values of 2.5 standard deviations or more away from the participant's mean of each vowel separately were removed in order to exclude any mistakes originated by the automatic alignment. After the outlier removal, the dataset included 7,977 words containing /æ/ and 7,648 words containing /ɛ/ in the pretest and 8,187 /æ/ tokens and 7,857 /ɛ/ tokens in the posttest.

The formant values were normalized using the Lobanov transformation, a vowelextrinsic normalization procedure which has been proven to be an efficient method to remove irrelevant physiological information about the speaker while preserving the phonemic variation (Adank et al., 2004).

3.4.3. Formant Analyses

3.4.3.1. Analysis of the pre-recorded speech

First, we analyzed the values of the pre-recorded speech used for both settings to confirm the two vowels were distinguished in terms of height and frontness in the accent of the speaker who recorded the speech samples. In order to do so, the audio samples played during the interaction were forced aligned and the formants were extracted using the same procedure as in 3.2. The data were then analyzed with linear mixed effect models fitted to the F1 and the F2 independently using the *lmerTest* package (version 3.1.0) (Kuznetsova et al., 2017) in R (version 3.5.0) (R Development Core Team, 2008). The fixed effect structure included the phoneme type (either /æ/ or /ɛ/) as a fixed factor. The model also included a random intercept for Word. The output of the F1 model with the /æ/ vowel on the intercept revealed a significant effect of Phoneme (β=-127, $p = 0.001$), indicating that /æ/ had indeed a higher F1 value than /ɛ/ words. Therefore, this speaker produced /æ/ tokens lower than /ɛ/ tokens. Similarly, the output of the F2 model also revealed a significant effect of Phoneme (β=344.17, $p = <0.001$), which means that the two vowels are also distinguished in terms of frontness: /æ/ tokens are significantly less fronted than /ɛ/ tokens. Interestingly, the effect size for Phoneme was bigger in the F2 model than in the F1 model, suggesting that, for this speaker, the difference between the vowels in terms of frontness was bigger than in terms of height.

3.4.3.2. Participants' speech analysis

We analyzed the two formants separately as we expected that, since Dutch speakers have been found to show asymmetrical knowledge of the two dimensions (Troncoso-Ruiz et al., 2019), the manipulations of the study would affect the F1 and the F2 dimensions differently. Dutch learners of English have shown to be able to distinguish the English /æ-ɛ/ vowels in terms of height but not in terms of frontness. In addition, previous research has found that non-native speakers can show differential learning for the two formants (Wanrooij et al., 2013).

In order to determine whether participants enlarged the distinction between the two vowels in terms of height and frontness, linear mixed effect models were fitted to the transformed data of the two formants independently using the *lmerTest* package (version 3.1.0) (Kuznetsova et al., 2017) in R (version 3.5.0) (R Development Core Team, 2008).[1] Methods that combine the two formants in a single outcome variable, for instance Markov Chain Monte Carlo Generalized Linear Mixed Models or Euclidean distances, were deemed less suitable as they would hide the differences between the two formants.

The focus of this study is to find out whether the difference between the two vowels in terms of F1 and F2 values significantly increased from pre-test to post-test depending on the condition and the setting to which participants had been assigned. Thus, the fixed-effect structure included as predictors: Vowel (/æ/ or /ɛ/), Test (Pre/Post), Condition (Control, OnlyExposure, OnlyFeedback or ExposureAndFeedback), and Setting (Human-human or Human-computer). If the participants distinguished these two vowels differently after the interaction depending on the setting and the condition to which they had been assigned, the interaction between Test, Vowel, Condition and Setting would be expected to be significant. As for the random effect structure, the model included random intercepts for Speaker and Word, to ensure that any differences found were not driven by just a few speakers or by uncontrolled properties of the words. We also included a random slope for Speaker by Test, so that the differences among the conditions were not driven by individual participants. No random slopes for Word were included due to non-convergence issues. The fixed-effect structure was improved by removing the interactions and simple effects with p-values higher than 0.05 that were not part of other significant interactions, but only when removal improved the model AIC value. The final models were refit removing the outliers of the model (2.5 standard deviations) to

---

[1] The data regarding the human-human setting was also analyzed separately in Troncoso-Ruiz et al. (2019).

confirm the significant effects and interactions were not relying on outliers.

Table 1 Analysis of Deviance Table (Type II Wald chi-square tests) for the Lobanov-transformed F1 and F2 values

Response: Lobanov-transformed F1 values			
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.02	1	0.882
Condition	2.47	3	0.480
Vowel	**37.73**	**1**	**<0.001**
Setting	**383.46**	**1**	**<0.001**
Test:Condition	3.52	3	0.319
Test:Vowel	**11.58**	**1**	**<0.001**
Test:Setting	0.04	1	0.835
Condition:Vowel	**35.35**	**3**	**<0.001**
Condition:Setting	4.27	3	0.234
Vowel:Setting	**269.06**	**1**	**<0.001**
Test:Condition:Vowel	**29.17**	**3**	**<0.001**
Test:Condition:Setting	2.13	3	0.545
Test:Vowel:Setting	0.34	1	0.562
Condition:Vowel:Setting	**61.39**	**3**	**<0.001**
Test:Condition:Vowel:Setting	**14.65**	**3**	**0.002**
Response: Lobanov-trasnformed F2 values			
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.05	1	0.821
Condition	5.87	3	0.118
Vowel	**6.04**	**1**	**0.014**
Setting	**65.31**	**1**	**<0.001**
Test:Condition	0.98	3	0.806
Test:Vowel	**19.26**	**1**	**<0.001**
Test:Setting	0.04	1	0.846
Condition:Vowel	**57.74**	**3**	**<0.001**
Condition:Setting	3.06	3	0.382
Vowel:Setting	0.78	1	0.375
Test:Condition:Vowel	**24.42**	**3**	**<0.001**
Test:Condition:Setting	1.17	3	0.760
Test:Vowel:Setting	**21.40**	**1**	**<0.001**
Condition:Vowel:Setting	**64.88**	**3**	**<0.001**
Test:Condition:Vowel:Setting	**123.21**	**3**	**<0.001**

Table 1 includes the analysis of deviance tables for the models fitted to the Lobanov-transformed F1 and F2. This table was obtained using the "Anova" function of the car package (version 3.0.2) (Weisberg & Fox, 2011) in R (version 3.5.0) (R Development Core Team, 2008). This function calculates the Type II analysis-of-variance table for the regression models. The outputs of the two models show significant interactions between Test, Condition, Vowel and Setting. These interactions indicate that speakers did indeed adapt the height and frontness of the critical vowels differently depending on the combination of the setting (human-human and human-computer) and the condition to which they had been assigned. In order to interpret this 4-way interaction, we split the data by condition and analyzed the subsets separately comparing the difference settings. The following sections include the analyses per condition comparing the two settings.

3.4.3.3. Control Condition (baseline condition)

In the Control condition, participants received no exposure to the critical vowels or feedback during the interaction with the confederate, human or computer.

Figure 4 shows the distribution of the F1 and F2 values of the two vowels as produced by the participants in the pre-test and the post-test in this condition. The figure suggests that in the human-human setting, participants did not adapt the pronunciation of neither of the vowels in neither of the dimensions, height and frontness. In the human-computer setting, on the other hand, the overlap between the two vowels seems to be different from the pre-test to the post-test. In particular, participants in this setting appear to be adapting both the height and frontness. This observation is supported by the statistical analyses.

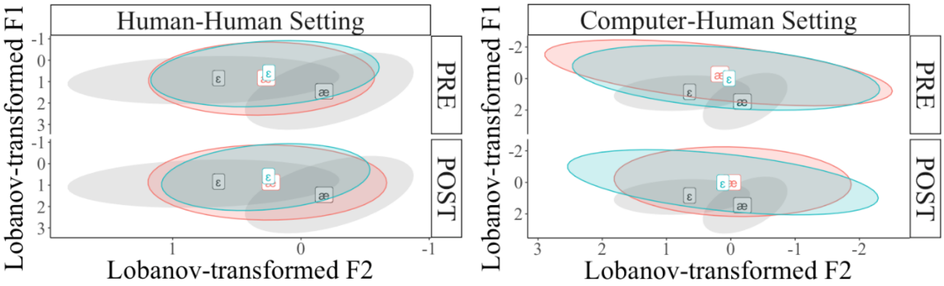


Figure 4. Vowel plots with ellipses representing the distribution of the F1 and F2 values of the /æ/and /ɛ/ when produced in the pre-test (top) and post-test (bottom) by participants in the human-human (left) and the human-computer (right) settings for the Control condition. The ellipses in grey represented the pronunciation of the same vowels by the native speaker who was recorded for the pre-recorded utterances.

Table 2 shows the outputs of the analysis of deviance of the models fitted to the F1 and F2 data from the participants in the Control condition. Both the F1 and the F2 models show a significant interaction between Test, Vowel and Setting, indicating that the difference between two vowels changed from pre-test to post-test differently depending on the setting. In order to understand how the difference between the two vowels was affected from pre-test to post-test in the two settings, the data was analyzed split by Setting.

For the human-human setting, the output of the models fitted to the F1 and F2 values showed that neither the simple effect of Test (F1: β = 0.04, t = 1.185, p = 0.248; F2: β =-0.02, t = -0.444, p = 0.664) nor the interaction between Test and Vowel (F1: β = -0.07, t = -1.456, p = 0.145; F2: β = 0.01, t = 0.469, p = 0.639) were significant (see Appendix 4). Thus, no evidence was found for the distinction between the two vowels to be different from pre-test to post-test in the Control condition of the human-human setting.

As for the human-computer setting, both models show the interaction between Test and Vowel to be statistically significant (F1: β = -0.26, t= -4.299, p=<0.001; F2: β = 0.27, t = 4.363, p=<0.001) (see Appendix 5). These interactions indicate that the distinction between the vowels in terms of height and frontness changed from pre-test to post-test in the Control condition of the human-computer setting. In order to understand how the distinction was affected, we take into account the effect of Vowel and how it was modulated by Test. In terms of height, the model with the pre-test productions of the /æ/ tokens on the intercept revealed that the F1 values of the vowel on the intercept, /æ/, were significantly lower than those of /ɛ/ in the pre-test (Vowel: β=0.27, t = 6.363, p = <0.001) (see Appendix 5). Therefore, participants distinguished these two vowels in the pre-test opposite to the native-like way. The interaction between Test and Vowel in this model indicates that the distinction between the vowels changed from pre-test to post-test; in order to understand this change, the model was relevelled to include the posttest productions on the intercept (see Appendix 6). The output of the model revealed no significant effect of Vowel in the post-test, suggesting the F1 values of the two vowels were not found to be significantly different in the post-test. Thus, participants' distinction of the two vowels became less accented in the post-test as compared to the pre-test. In order to understand whether this pretest-posttest change was driven specifically by the F1 values of one of the two vowels varying significantly from pre-test to post-test, the simple effects of Tests were taken into account. The model with the pretest productions of /æ/ on the intercept revealed a significant effect of Test (β=0.28, t=2.214 , p=0.045) (see Appendix 5), suggesting that the F1 values of this vowel were significantly higher in the post-test than in the pre-test. Relevelling the model to include the /ɛ/ vowel on the intercept revealed no significant effect of Test for this vowel (see Appendix 7).

As for F2, the model with the pre-test productions of the /æ/ tokens on the intercept revealed a significant effect of Vowel (β=-0.13, t = -2.991, p=0.003) (see Appendix 5), suggesting that the F2 values of the vowel on the intercept, /æ/, were significantly higher than the F2 values of /ɛ/ in the pre-test. This distinction also follows the opposite pattern to native-like way of distinguishing these two vowels. The interaction between Test and Vowel indicates that the distinction between the vowels changed from pre-test to post-test. Releveling the model to include the post-test tokens on the intercept revealed a significant effect of Vowel (β=0.14, t = 3.090 , p = 0.002) (see Appendix 6). This effect indicates that, in the post-test, the F2 values of the vowel on the intercept, /æ/, were lower than those of /ɛ/. The frontness distinction between the vowels in the posttest did follow the native-like pattern. In order to investigate whether this change was driven specifically by the F2 values of one of the vowels changing significantly from pretest to post-test, the simple effects of Test were taken into account. No significant effect of Test was found in the models with either /æ/ or /ɛ/ tokens in the intercept (see Appendix 5 and 7).

In summary, participants in the human-human setting showed no adaptation of their pronunciation of the vowels at all, neither in terms of height or frontness. Participants in the human-computer setting, on the other hand, adapted their pronunciation of the two vowels both in terms of height and frontness, displaying a more native-like distinction of the two vowels in the post-test than in the pre-test.

Table 2 Output from the analysis of deviance table including Type II Wald chisquare tests of the models fitted to the Lobanov-transformed formant values (F1 and F2) as predicted by Test, Vowel and Setting and their interactions. Values in bold are considered significant.

	Lobanov-transformed F1		
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	1.66	1	0.197
Vowel	**5.05**	**1**	**0.025**
Setting	**70.70**	**1**	**<0.001**
Test:Vowel	**15.58**	**1**	**<0.001**
Test:Setting	0.97	1	0.324
Vowel:Setting	**112.26**	**1**	**<0.001**
Test:Vowel:Setting	**6.33**	**1**	**0.012**
	Lobanov-transformed F2		
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.09	1	0.765
Vowel	0.40	1	0.526
Setting	**13.70**	**1**	**<0.001**
Test:Vowel	**17.87**	**1**	**<0.001**
Test:Setting	0.03	1	0.872
Vowel:Setting	0.06	1	0.798
Test:Vowel:Setting	**10.89**	**1**	**<0.001**

3.4.3.4. OnlyExposure Condition (RQ1)

In the OnlyExposure Condition, participants received exposure to the interlocutor's pronunciation of the critical contrast in the context of minimal pairs. Following previous research, we predicted participants interacting with the human confederate to show overall more phonetic accommodation than participants interacting with the computer.

In Figure 5, where the distribution of the two vowels is represented in the pretest and post-test for the two settings in the OnlyExposure Condition, participants appear to adapt their pronunciation differently in the two settings. Whereas there seems to be convergence to the native-like pronunciation in the human-human setting, the ellipses for the vowels in the human-computer setting seem to in fact show divergence. In the human-computer setting, participants seem to adapt the production of /ɛ/ in terms of frontness, producing less fronted tokens. These patterns are confirmed by the statistical analyses of the F1 and F2.

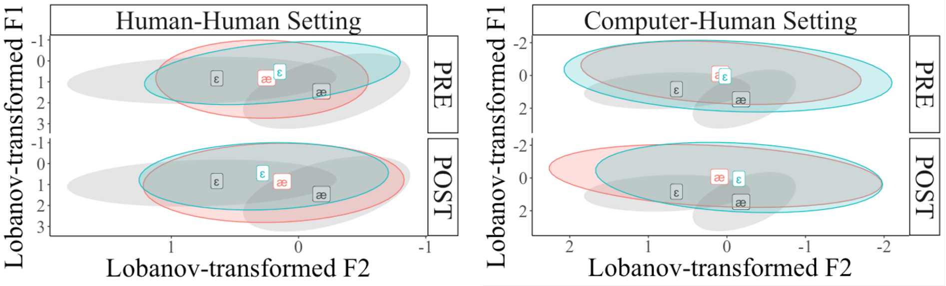


Figure 5. Vowel plots with ellipses representing the distribution of the F1 and F2 values of the /æ/ and /ɛ/ when produced in the pre-test and post-test by participants in the human-human (left) and the human-computer (right) settings for the Only Exposure Condition. The ellipses in grey reprsented the pronunciation of the same vowels by the native speaker who was recorded for the pre-recorded utterances.

Table 3 shows the analysis of deviance of the models fitted to the F1 and F2 data of the participants in the OnlyExposure Condition. The F1 model showed no significant effect of Test and no significant interactions including Test either. These results indicate that in the OnlyExposure condition the height dimension remained unchanged from pre-test to post-test for the two vowels regardless of whether participants interacted with a human or a computer. As for F2, the model revealed a significant interaction between Test, Vowel and Setting. This interaction suggests that the distinction between the two vowels was affected differently from pre-test to post-test depending on the setting. In order to understand how the F2 values were affected in the two settings, the F2 data were split by Setting.

For the human-human setting, the model including the pre-test productions of /æ/ on the intercept showed a significant interaction between Test and Vowel (β=0.23, t=7.811 , p=<0.001) (see Appendix

8), indicating that the distinction in F2 between the two vowels was significantly different from pre-test to post-test. With the pre-test productions of /æ/ on the intercept, the simple effect of Vowel was not found to be significant (β=-0.03, t=-0.713, p=0.5) (see Appendix 8); thus, no differences were found between the F2 values of the two vowels in the pre-test. Relevelling the model to place the post-test productions on the intercept revealed a significant simple effect of Vowel (β=0.20, t=5.255, p=<0.001) (see Appendix 9), indicating that the F2 values of the vowel /æ/, placed on the intercept, were significantly lower than those of /ɛ/ in the post-test. This distinction follows the native-like way of distinguishing these two vowels in terms of frontness. In order to understand whether this pretest-posttest change was driven specifically by the F2 values of one of the vowels significantly changing from pre-test to post-test, we examined the simple effect of Test. The model with the pre-test productions of /æ/ on the intercept revealed a significant effect of Test (β=-0.11, t=-2.557, p=0.023) (see Appendix 8), which indicates that the F2 values of this vowel were significantly lower in the post-test as compared to the pre-test. Relevelling the model to include the /ɛ/ vowel on the intercept also revealed a significant simple effect of Test (β=0.12, t=-7.811, p=<0.001) (see Appendix 9); the F2 values of /ɛ/ were found to be significantly higher in the post-test than in the pre-test.

In the human-computer setting, the model with the pre-test productions of /æ/ tokens on the intercept revealed a significant interaction between Test and Vowel (β=-0.18, t=3.038 , p=0.002) (see Appendix 10), indicating that the distinction between the two vowels in terms of frontness changed from pre-test to post-test in the human-computer setting. This model also revealed that the simple effect of Vowel (β=-0.00, t=-0.012, p=0.991) was not significant for the pre-test productions; thus, no evidence was found for the F2 values of the two vowels to be significantly different in the pre-test. In order to understand how the Vowel effect was different in the post-test, the model was relevelled to include the post-test productions on the intercept (see Appendix 11). The output of the relevelled model revealed a significant effect of Vowel (β=-0.25, t=-6.192, p=<0.001) for the post-test productions. This effect confirms that, in the post-test, the vowel on the intercept, /æ/, was produced with significantly higher F2 values than /ɛ/. This distinction between the vowels represents the opposite of the native-like distinction between the vowels. In order to understand whether this pretest-posttest change was driven specifically by the F2 values of one of the vowels varying significantly from pretest to post-test, the imple effect of Test was taken into account. No simple effects of Test were found in neither the model with the pre-test tokens of /æ/ on the intercept (see Appendix 10) nor in the relevelled model with the pre-test /ɛ/ tokens on the intercept (see Appendix 12). Thus, the distinction between the two vowels changed from pre-test to post-test to become less native-like, but neither of the two vowels was found to be produced with F2 significantly different from pre-test to post-test independently.

In summary, participants in the OnlyExposure condition showed differences depending on whether they interacted with the human confederate or the computer. In the case of the human-human setting, participants produced more native-like vowels in the post-test than in the pre-test by enlarging the distance between the two vowels in terms of frontness in the native-like way. Participants in the human-computer setting, on the other hand, showed changes in the frontness too, but following the direction opposite to what is expected of the native-like distinction of the two vowels. The height dimension remained unchanged in both settings in this condition.

Table 3 Output from the models fit to the Lobanov-transformed formant values (F1 and F2) as predicted by Test, Vowel and Setting and their interactions. The model includes in the pre-test, the human-human setting and /æ/ vowel. Values in bold are considered significant.

	Lobanov-transformed F1		
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.21	1	0.644
Vowel	**13.71**	**1**	**<0.001**
Setting	**70.68**	**1**	**<0.001**
Test:Vowel	3.31	1	0.069
Test:Setting	0.00	1	0.970
Vowel:Setting	**156.60**	**1**	**<0.001**
Test:Vowel:Setting	0.62	1	0.432
	Lobanov-transformed F2		
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.69	1	0.406
Vowel	**4.11**	**1**	**0.043**
Setting	**8.55**	**1**	**0.003**
Test:Vowel	2.92	1	0.088
Test:Setting	0.06	1	0.433
Vowel:Setting	**37.58**	**1**	**<0.001**
Test:Vowel:Setting	**45.14**	**1**	**<0.001**

3.4.3.5. OnlyFeedback Condition (RQ2)

In the OnlyFeedback Condition, participants received feedback suggesting that the interlocutor always misunderstood their pronunciation of /æ/ as /ɛ/ during the interaction, but they never received exposure to the native pronunciation of the contrast. We expected differences in how participants adapted their pronunciation depending on whether they thought the feedback was coming from a computer or a human interlocutor.

Figure 6 shows that participants in the OnlyFeedback condition who interacted with the human confederate appear to produce more native-like tokens after the interactive task: increasing the height and frontness distinction between the two vowels from the pre-test to the post-test. Participants in the human-computer setting appear to also increase the height distinction of the tokens in a native-like way after the interactive task, but they also show adjustments to the frontness dimension in the direction opposed to the native-like distinction. In particular, they seem to produce more fronted /æ/ tokens in the post-test as compared to the pre-test. This is confirmed by the statistical analyses.

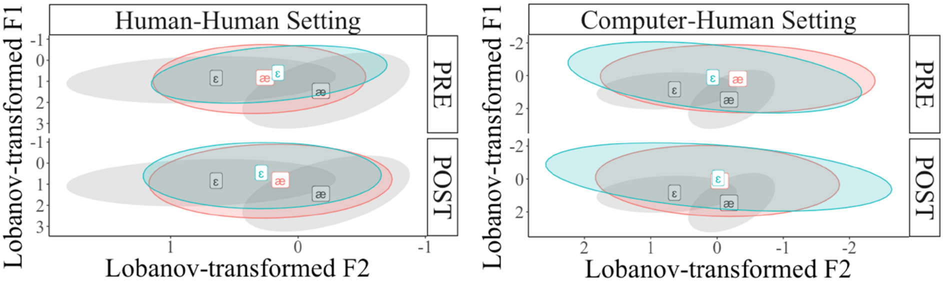


Figure 6. Vowel plots with ellipses representing the distribution of the F1 and F2 values of the /æ/ and /ɛ/ when produced in the pre-test and post-test by participants in the human-human (left) and the human-computer (right) settings for the Only Feedback Condition. The ellipses in grey represented the pronunciation of the same vowels by the native speaker who was recorded for the pre-recorded utterances.

Table 4 includes the analysis of deviance of the models fitted to the F1 and F2 values of the participants in the OnlyFeedback Condition. The interaction between Test, Vowel and Setting was not found to be significant for the F1 model. No evidence was found, thus, for the F1 values to be affected differently in the two settings. The F1 model revealed a significant interaction between Test and Vowel, indicating that the distinction between the two vowels changed from pre-test to post-test similarly in the two settings.

With the pre-test /æ/ tokens produced by the participants in the human-human setting on the intercept, the model revealed a significant effect of Vowel (β=-0.22, t=-5.040, p=<0.001) (see Appendix 13), confirming that the F1 values of the /æ/ tokens were higher than those of /ɛ/ in the pre-test. This distinction follows the native-like pattern to distinguish these two vowels. In order to understand how the distinction changed from pre-test to post-test, the model was relevelled to include the post-test productions on the intercept (see Appendix 14). The output of this relevelled model also revealed a significant effect of Vowel (β=-0.35, t=-7.802, p=<0.001), with a larger beta. Thus, in the post-test, participants increased the F1 distinction between the two vowels following the native-like pattern. In order to investigate whether this pretest-posttest difference was driven in particular by the F1 of one of the vowels changing significantly, the simple effect of Test was taken into account. The effect of Test was not found to be significant in either the model with the /æ/ vowel on the intercept (see Appendix 13) nor the model with the /ɛ/ tokens on the intercept (see Appendix 15)

As for the frontness dimension, the model (see Table 4) revealed a significant interaction between Test, Vowel and Setting, suggesting that the distinction between the vowels varied from pre-test to post-test differently depending on the setting to which participants had been assigned. In order to interpret how the difference between the vowels was affected in each setting, the data was split by Setting.

The model fitted to the F2 data from the human-human setting revealed a significant interaction between Test and Vowel (β=0.23, t=8.385, p=<0.001) (see Appendix 16), indicating that the distinction between the two vowels in frontness changed from pre-test to post-test. In particular, the model with the pre-test production of the /æ/ tokens on the intercept showed no significant simple effect of Vowel (β=-0.03, t=-0.836, p=0.404); thus, no evidence was found for the F2 values of the two vowels to be significantly different in the pre-test. In order to understand the interaction between the Test and Vowel, the model was relevelled to include the post-test productions on the intercept instead (see Appendix 17). The output of this relevelled model showed a significant effect of Vowel (β=0.20, t=5.373, p=<0.001) for the post-test productions, confirming that, in the post-test, the F2 values of the vowel on the intercept, /æ/, were significantly lower than the F2 values of /ɛ/. This distinction follows the native-like pattern to distinguish these two vowels. In order to understand whether this pretest-posttest change was driven specifically by one of the vowels in particular, the simple effect of Test was taken into account. The model with the pre-test /æ/ tokens on the intercept revealed a significant effect of Test (β=-0.11, t= -3.433, p=0.003) (see Appendix 16) for this vowel, indicating that the F2 values of this vowel were found to be significantly lower in the post-test than in the pre-test. Similarly, with the /ɛ/ vowel on the intercept, the simple effect of Test was also found to be significant (β=0.11, t=3.282, p=0.004) (see Appendix 18). In this case, however, the direction of the change was the opposite: the F2 values of /ɛ/ were found to be higher in the post-test than in the pre-test.

The model fitted to the F2 data from the human-computer setting also revealed a significant interaction between Test and Vowel (β=-0.35, t=-5.495, p=<0.001) (see Appendix 19), confirming that the difference in F2 between the vowels changed from pre-test to post-test. The significant effect of Vowel (β= 0.38, t=8.180, p=<0.001) indicates that, in the pre-test, the F2 values of the vowel on the intercept, /æ/, were significantly lower than the F2 values of the /ɛ/ tokens. This distinction followed the native-like pattern to distinguish these two vowels. In order to understand how this effect had been modulated by Test, the model was releveled to include the post-test productions on the intercept instead (see Appendix 20). The relevelled model revealed no significant effect of Vowel (β=0.03, t=0.718, p=0.473); thus, no evidence was found for F2 values of the two vowels to be different in the post-test. In order to understand whether this pretest-posttest change was driven particularly by one of the vowels, the simple effect of Test was taken into account. The model including the pre-test /æ/ tokens on the intercept revealed a significant effect of Test (β=0.26, t=2.749, p=0.016) (see Appendix 19), confirming that the F2 values of the vowel on the intercept were significantly higher in the post-test than in the pre-test. Relevelling the model to include the /ɛ/ on the intercept revealed no significant effect of Test (β=-0.08, t=-0.847, p=0.412) (see Appendix 21); thus, no evidence was found for the F2 values of this vowel to be different from pre-test to post-test.

In brief, whereas the height distinction between the two vowels was increased from pre-test to post-test similarly in the two settings; the frontness distinction was affected differently depending on the setting. Only participants in the human-human setting improved the difference between these two vowels in the native-like way in the post-test. In the human-computer setting, on the other hand, the participants showed adjustments in the difference between the two vowels going in the direction opposite to the native-like pattern to distinguish these vowels.

Table 4 Output from the models fit to the Lobanov-transformed formant values (F1 and F2) as predicted by Test, Vowel and Setting and their interactions. The model includes on the intercept the pre-test, the human-human setting and /æ/ vowel. Values in bold are considered significant.

			Lobanov-transformed F1
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	1.15	1	0.283
Vowel	**45.57**	**1**	**<0.001**
Setting	**100.17**	**1**	**<0.001**
Test:Vowel	**10.52**	**1**	**0.001**
Test:Setting	0.10	1	0.746
Vowel:Setting	**15.85**	**1**	**<0.001**
Test:Vowel:Setting	0.01	1	0.902
			Lobanov-transformed F2
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.83	1	0.363
Vowel	**31.79**	**1**	**<0.001**
Setting	**43.91**	**1**	**<0.001**
Test:Vowel	0.95	1	0.329
Test:Setting	1.14	1	0.286
Vowel:Setting	**29.41**	**1**	**<0.001**
Test:Vowel:Setting	**77.26**	**1**	**<0.001**

3.4.3.6. ExposureAndFeedback Condition (RQ3)

Participants in the ExposureAndFeedback condition received feedback on their pronunciation in addition to exposure to the native pronunciation of the critical contrast in the context of a minimal pair. We expected the effect of these two factors in combination to differ in the human-human and human-computer settings. Following previous research (Burnham et al., 2010), we expected speakers to show more adaptations in the human-computer setting than in the human-human setting.

Figure 7 shows the distribution of the two vowels produced in the pre-test and post-test in the two settings in the ExposureAndFeedback Condition. This figure shows that participants in the two settings reacted differently to the combination of the two factors. Participants in the human-human setting show a similar degree of overlap between the vowels in the pre-test and post-test, suggesting no adaptation at all. In the human-computer setting, however, the overlap seems to differ drastically between the pre-test and the post-test. In particular, participants seem to produce higher and more back /æ/ tokens in the post-test than in the pre-test. We observed the same patterns in the statistical analyses.

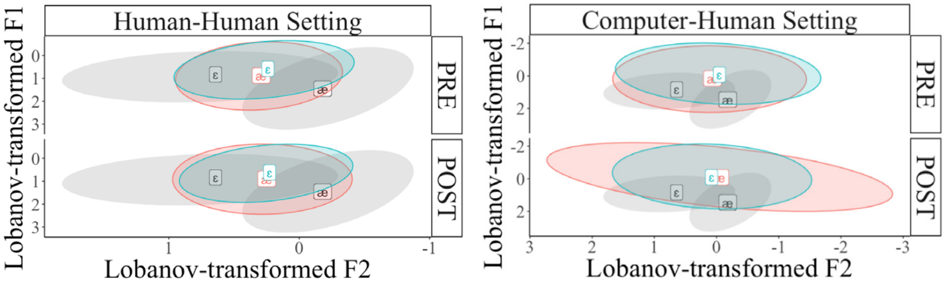


Figure 7. Vowel plots with ellipses representing the distribution of the F1 and F2 values of the /æ/ and /ɛ/ when produced in the pre-test and post-test by participants in the human-human (left) and the human-computer (right) settings for the ExposureAndFeedback condition, with exposure and feedback combined. The ellipses in grey represented the pronunciation of the same vowels by the native speaker who was recorded for the pre-recorded utterances.

Table 5 shows the analysis of deviance of the F1 and F2 models fitted to the data from the participants in ExposureAndFeedback Condition. The interaction between Test, Vowel and Setting is significant in both models. These significant interactions indicate that, depending on the setting, the difference between the two vowels in the pre-test and post-test varied. In order to interpret how the distinction changed from pre-test to post-test in the two settings, the data was split by Setting.

The models fitted to the F1 and the F2 values from the participants in the human-human setting revealed no significant interactions between Test and Vowel (F1: β=0.01, t=0.233, p=0.816; F2: β=0.03, t=1.612, p=0.107) (see Appendix 22). These models did not reveal a significant simple effect of Test either (F1: β=0.02, t=0.606, p=0.551; F2: β=-0.04, t=-1.469, p=0.165). Thus, no evidence was found for the distinction between the two vowels to be different from pre-test to post-test in this setting neither in terms of height or frontness, as illustrated in the Figure 7.

As for the human-computer setting, the model fitted to the F1 data with the pre-test productions of the /æ/ tokens on the intercept revealed a significant interaction between Test and Vowel (β=0.23, t=3.763, p=<0.001) (see Appendix 23). This interaction indicated that the difference between the two vowels in terms of F1 changed from pre-test to post-test. The model revealed a significant simple effect of Vowel (β=-0.24, t=-5.488, p=<0.001), which indicated that, in the pre-test productions, the F1 values of the vowel on the intercept, /æ/, were significantly higher than the values of the /ɛ/ tokens. This distinction follows the native-like pattern to discriminate these two vowels. In order to understand how the distinction was affected from pre-test to post-test, the model was relevelled to include the post-test productions on the intercept (see Appendix 24). The output of the model revealed no significant simple effect of Vowel (β=-0.02, t=-0.433, p=0.664); thus, no evidence was found for the F1 values of the two vowels to be different in the post-test. In order to understand whether this change, from a larger distinction of the vowels in the pre-test to a smaller distinction between the vowels on the post-test, was driven in particular by one of the vowels, the simple effect of Test was taken into account. No significant effect of Test was found either in the model with the pre-test productions of /æ/ tokens on the intercept (see Appendix 23), neither in the model with the pre-test productions of /ɛ/ on the intercept (see Appendix 28)

As for the F2, the model with the pre-test productions of /æ/ tokens on the intercept revealed a significant interaction between Test and Vowel (β=0.25, t=4.166, p=<0.001) (see Appendix 23), indicating that the distinction between the vowels was different in the pre-test and in the post-test. This model also revealed no significant simple effect of Vowel; thus, no evidence was found for the F2 values of the two vowels to be significantly different in the pre-test. In order to understand how the distinction between the two vowels changed from pre-test to post-test, the model was relevelled to include the post-test productions on the intercept (see Appendix 24). The relevelled model included a significant effect of Vowel ($\beta=0.17$, $t=3.707$, $p=<0.001$), which indicates that, in the post-test, the F2 values of the vowel on the intercept, /æ/, were significantly lower than those of /ɛ/. This distinction follows the native-like pattern to distinguish these two vowels. In order to understand whether this change was driven by the F2 values of one of the vowels in particular varying from pre-test to post-test, the simple effect of Test was taken into account. No significant effect of Test was found in neither the model including the pre-test productions of /æ/ on the intercept (see Appendix 23) nor in the relevelled model including the pre-test /ɛ/ tokens on the intercept (see Appendix 25).

In summary, participants in this condition also showed differences in phonetic accommodation depending on the setting. While participants in the human-human setting did not adapt their pronunciation of any of the two vowels from pre-test to posttest, neither in terms of height or frontness; participants who interacted with the computer adapted their pronunciation of the vowels: they distinguished the two vowels in a more native-like way in terms of frontness but produced less native-like tokens in terms of height in the post-test.

Table 5 Output from the models fit to the Lobanov-transformed formant values (F1 and F2) as predicted by Test, Vowel and Setting and their interactions. The model includes on the intercept the presence of feedback, the presence of exposure, the pre-test, the human-human setting and æ vowel. Values in bold are considered significant.

			Lobanov-transformed F1
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.49	1	0.486
Vowel	**29.48**	**1**	**<0.001**
Setting	**388.66**	**1**	**<0.001**
Test:Vowel	**11.81**	**1**	**<0.001**
Test:Setting	1.06	1	0.304
Vowel:Setting	**23.06**	**1**	**<0.001**
Test:Vowel:Setting	**9.50**	**1**	**0.002**
			Lobanov-transformed F2
Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	0.04	1	0.835
Vowel	2.78	1	0.095
Setting	**27.95**	**1**	**<0.001**
Test:Vowel	**21.08**	**1**	**<0.001**
Test:Setting	0.17	1	0.683
Vowel:Setting	2.17	1	0.141
Test:Vowel:Setting	**12.37**	**1**	**<0.001**

3.5. General discussion

The current study investigates whether speakers adapt their pronunciation differently when they interact with a human interlocutor or with a computer. In particular, we assess how phonetic accommodation in these two settings was affected by exposure to the native-like pronunciation (RQ1), feedback on the speaker's pronunciation (RQ2) and the combination of exposure and feedback (RQ3). Results show that, overall, participants adjusted their pronunciation of the two critical vowels differently depending on whether the interlocutor was presented as a human or a computer interlocutor. The different factors under study also triggered different degrees of phonetic accommodation. Exposure, in isolation, triggered native-like phonetic accommodation from the participants who interacted with the human interlocutor, but not from the participants who interacted with the computer interlocutor. Similarly, feedback on the pronunciation was overall more effective in fostering native-like phonetic accommodation in the human context than in the human-computer setting. Surprisingly, when the two factors, exposure and feedback, were presented in combination during the interaction, participants in the human-computer setting showed more accommodation than the ones who interacted with the human confederate, who showed no adjustments at all.

Overall, the vowel /æ/ seemed to be more affected by the effects of the exposure and feedback than its counterpart /ɛ/. This asymmetry between the vowels could be related to the difference in space for improvement these two vowels have when produced in a Dutch-accented way. Since Dutch speakers tend to merge these two vowels into an in between sound that is closer to English /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006), /æ/ may have more room for improvement and, therefore, may be more affected than /ɛ/. This idea falls in line with previous research investigating the role of perceptual salience, i.e., perceptual distance, in phonetic accommodation (MacLeod, 2014). This literature proposes that phonetic features that are more perceptually different are more likely to be accommodated. Therefore, we could expect the Dutch-accented vowel that is acoustically more different from the native pronunciation to be more likely affected by phonetic accommodation. In addition, another aspect that could contribute to /æ/ undergoing more changes than /ɛ/ is the two different roles these vowels had during the puzzle-solving game. In all the critical Code Breaker trials, participants had to say the word with the /æ/ vowel, and never the /ɛ/ word. This prominent position during the game may have also contributed to the overall larger change this vowel underwent. This reasoning is supported by the fact that we see this pattern of /æ/ undergoing more changes even in the baseline condition, where participants' role involved producing the /æ/ vowel without receiving exposure to the vowel contrast or feedback on their pronunciation.

The results of this study showed that the two dimensions that distinguish the English contrast /æ-ɛ/, namely height and frontness, were affected differently when speakers adapted their pronunciation. This finding falls in line with previous literature investigating the learning of L2 contrasts (for instance, see Wanrooij et al., 2013). Non-native learners of a language have been found to show asymmetrical learning of the two formant dimensions. One possible explanation could be related to the baseline knowledge of these two dimensions participants had before the interaction. Dutch speakers of English have been found to master these two dimensions at different levels (Troncoso-Ruiz et al., 2019); in particular, they were found to produce significantly different values for the F2 of these two vowels, but not for the F1 dimension. This asymmetry suggests that Dutch speakers can distinguish these two vowels in terms of frontness but not in terms of height. Therefore, difference in room for improvement in the two formant dimensions could lead to an asymmetrical learning process.

3.5.1. Control condition

Participants in the human-computer setting who did not receive exposure or feedback showed a change in their pronunciation of the critical vowels, which suggests that only taking part in the interaction already triggered an adaptation. This adaptation could be related to the nature of the task: during the interactive game, participants were visually exposed to minimal pairs containing the critical contrast. Previous research, for instance Escudero-Mancebo et al. (2015), has shown that L2 learners can achieve a more native-like pronunciation of difficult L2 sounds after being visually (and optionally also auditorily) exposed to a battery of minimal pairs containing these sounds. In Escudero-Mancebo (2015), exposure to the pronunciation of the vowels was not fully controlled for: participants could choose to hear the vowels while performing the task, the model pronunciation and their own. In our study, participants also showed adaptation in the pronunciation of the vowels even when they were strictly only visually exposed to the minimal pairs.

This group of participants in our experiment, who did not experience exposure to the critical sounds or feedback, showed differences depending on the type of interlocutor with whom they interacted during the experiment. In particular, participants interacting with the computer interlocutor adapted their vowels whereas those interacting with the human interlocutor did not. This finding is in line with the finding of Burnham et al. (2010) that speakers adapted their pronunciation more when interacting with a computer than with a human. A possible explanation for this difference between the settings, proposed by Burnham et al. (2010), is related to our knowledge, as technology users, about the limitations of the systems computers use for speech recognition. When interacting with a human, speakers can assume that the interlocutor can adapt their expectations in order to understand our pronunciation. Speakers are likely to be aware that this assumption is not valid in the context of human-computer settings, since computers are not expected to be as flexible. As a consequence, speakers might show more adaptations in their pronunciation to make up for the limitations of the speech recognition systems used.

3.5.2. OnlyExposure condition

Our first research question (RQ1) tested whether exposure to the native pronunciation of a difficult L2 contrast triggered more phonetic accommodation in interactions with a human interlocutor than in interactions where participants believed they talked to a computer. Results show that the exposure to the critical sound as pronounced by a native speaker indeed triggered more native-like phonetic accommodation when participants believed their interlocutor was human than when they thought the voice belonged to a computer. This outcome is in line with results from previous research comparing human-human and human-computer phonetic accommodation in less interactive settings, for instance Cohn et al. (2019), and it confirms previous literature observing that exposure is indeed more relevant in human-human settings than in human-computer settings (Branigan et al. 2010). In fact, the degree to which speakers adapt their pronunciation is mediated by whether their conversation partner was presented as a human or a computer. Therefore, the difference between the results in each setting shows that phonetic accommodation is indeed socially mediated, as discussed in Cohn et al. (2019) and Raveh et al. (2019). In other words, if one of the goals of phonetic accommodation is reducing social distance, it makes sense that speakers would accommodate more to a human interlocutor than to a computer, and that the influence of a human interlocutors' pronunciation of specific sounds should affect accommodation more than a computer's pronunciation.

Unlike previous research investigating the effect of exposure on phonetic accommodation, participants in this study did not show an increase in their degree of accommodation to sound more similar to the interlocutor when they received the exposure to the native pronunciation of the critical sounds from the computer interlocutor. In fact, they produced less fronted /ɛ/ in the post-test, which could be interpreted as evidence for divergence, rather than convergence. One possible explanation for this lack of alignment could be related to the interactive nature of the task used in this study. Previous literature studying phonetic accommodation while also carefully controlling the phonetic exposure participants receive uses low demanding tasks, such as shadowing tasks (Cohn et al., 2019), naming tasks (Delvaux & Soquet, 2007) or heavily prescribed question-answer games (Gessinger et al. 2019; Burnham et al., 2010). Participants in low demanding tasks could be unconsciously encouraged to imitate the voice to which they were exposed, even when the experimenter did not give instructions to do so. The task in this study, on the contrary, resembles a spontaneous conversation where the interlocutors have a common goal: solving puzzles. Preserving the conversational goals during the experiment could emphasize the interactive component and, therefore, this encouragement for imitation may have been diluted.

Surprisingly, in the human-computer setting, participants in the Only Exposure Condition adapted their pronunciation less than those in the baseline condition, where they were not exposed to the interlocutor’s pronunciation of the /æ-ɛ/ contrast at all. One possible explanation is that the Only Exposure Condition was cognitively relatively more demanding, since participants also had to process the exposure to the contrast they received. As a consequence, they might have divided their attention into producing the sounds, in the context of a minimal pair as in the baseline condition, and perceiving the sounds as produced by the interlocutor in the Semantic Relation trials. The division of attention into perception and production has been found to hinder the learning process of non-native speakers (e.g., Baese-Berk, 2019).

This study, thus, shows that the controlled presence of exposure to critical sounds does not always guarantee an increase in phonetic accommodation. In fact, these findings provide evidence that the differences in how exposure can trigger accommodation are related to how human the interlocutor is presented. The lack of convergent accommodation to the computer interlocutor confirms that the type of task used to study accommodation can have relevant consequences on the degree of adaptation participants show. Therefore, these findings call for the implementation of more interactive settings to study phonetic accommodation.

3.5.3. OnlyFeedback Condition

Our second research questions (RQ2) explored whether feedback affected phonetic accommodation differently when speakers thought they were talking to a human confederate than when they thought they interacted with a computer. Similar to the effect of exposure, the presence of feedback was overall more effective in fostering phonetic accommodation in the setting with the human confederate. Although in both settings feedback equally triggered adjustments of the height of the /ɛ/ tokens, participants in the human confederate condition also adapted the frontness of the two vowels, distinguishing them more clearly in the post-test than in the pre-test. Participants in the human-computer setting did not show convergent accommodation in the frontness of the vowels but they actually diverged producing more fronted /æ/ tokens.

The overall effectiveness of feedback in both settings provides evidence for the audience-design view on phonetic accommodation: participants showed significant changes in their production when they noticed during the interaction that the interlocutor could not identify the vowel they produced. This shows that, even without exposure to the critical sound, speakers can adjust their pronunciation of a difficult sound to ensure understanding. However, the fact that feedback triggered more convergent accommodation from speakers who interacted with the human interlocutor than from those who interacted with the computer interlocutor is arguably unexpected. Previous literature investigating phonetic accommodation in human-human and humancomputer interactions has found that ensuring communicative success is a key element in human-computer interactions, more than in human-human interactions (Branigan et al. 2010). Thus, one could expect for feedback implying misunderstanding to be more successful triggering phonetic accommodation in human-computer settings than in human-human settings. However, in this study participants who interacted with the human interlocutor adapted both height and frontness distinction of the two vowels in a native-like way, whereas those in the human-computer setting failed to adapt the frontness distinction in a native-like way. One possible explanation for this finding could be related to the control for exposure. Previous literature discussing the differences in human-human and human-computer settings involving the presence of feedback implying misunderstanding tend to present exposure to the sounds subject to the feedback in the same conversation. However, in this condition of the experiment, participants did not receive exposure to the contrast on which they were receiving feedback.

Interestingly, speakers in the human-computer setting showed less accommodation of their pronunciation in the Only Feedback condition than in the baseline condition, where they did not receive feedback about the intelligibility of their pronunciation of the contrast. A possible explanation could be related to the nature of the feedback. In this experiment, feedback was consistently negative, therefore, regardless of the actual pronunciation, participants always received negative feedback suggesting that the interlocutor misunderstood them. If participants interpreted this as the computer's inability to adapt to their pronunciation, as a consequence, they may have stopped trying to adapt their speech.

The fact that feedback triggered adaptation of the vowel height similarly in human-human and human-computer settings partly supports the validation of the CALL systems which aim to substitute L2 pronunciation human trainers with computer approaches. However, given that participants also showed adaptation in terms of frontness in the human-human but not the human-computer setting, feedback seems to be, in general, more effective at fostering phonetic accommodation when it comes from a human interlocutor.

These findings show that feedback under the exact same circumstances is interpreted differently depending on whether participants think it is coming from a human or a computer interlocutor. The fact that in this study feedback triggered more accommodation in the human-human setting than in the human-computer setting, suggests that factors, such as feedback, differ in how they affect phonetic accommodation depending on the linguistic setting where this phenomenon can take place. The differences found in the human-computer and the human-human setting in this study highlight the limitations of the generalizability of computer-based approaches to real-life interactions.

3.5.4. ExposureAndFeedback condition

Our third research question (RQ3) investigated whether the combination of the two factors in an interaction would trigger different degrees of phonetic accommodation in a human-human setting than in a human-computer setting. Surprisingly, the effects of exposure and feedback on phonetic accommodation in isolation differed drastically from when the two factors were presented in combination in both settings. Although in the human-human setting these factors in isolation had triggered phonetic accommodation, participants in the ExposureAndFeedback condition did not show any changes from pre-test to post-test in any of the vowels. In the human-computer setting, on the other hand, participants showed an overall adjustment to match the native pronunciation of the contrast in terms of frontness, even though both factors had not triggered phonetic accommodation in isolation.

Unlike the results from the Only Feedback Condition, the pattern in this condition falls in line with previous studies that have shown that, in situations of miscommunications, speakers can show more accommodation when interacting with computers than with other human confederates (Branigan, et al., 2010; Burnham et al., 2010). One possible reason could be related to the presence of exposure in the interaction together with feedback. In Burnham et al. participants received feedback implying misunderstanding of the corner vowels (/i, u, a/), while they also receive exposure to the interlocutor's pronunciation of these vowels. While in the Only Feedback Condition participants did not receive exposure to the native pronunciation of the /æ-ɛ/ contrast, in the ExposureAndFeedback Condition they could hear 12 tokens of each vowel during the interaction. Therefore, the presence of exposure to the critical sounds could have made this last condition more comparable to Burnham et al. than the Only Feedback Condition.

Surprisingly, in the human-human setting participants showed less phonetic accommodation when feedback was combined with exposure than when it was presented in isolation. One possible explanation could be related to how participants interpreted the combination of the two factors. Participants could have found the combination of the two factors misleading when it came from a human interlocutor. If they adjusted their pronunciation during the interaction as a result of exposure but at the same time also received feedback implying this effort was not good enough for the confederate to identify the vowel, participants could have been confused and could have stopped trying to distinguish the two vowels.

In the human-computer setting, on the other hand, the combination of the two factors had the opposite effect: these factors together triggered phonetic accommodation to the computer. One possible explanation for the asymmetry between the effect of feedback in isolation and in combination with exposure in the humancomputer setting could be explained by exposure only becoming relevant for the participants when they were also made aware, through feedback, of the difference between the nonnative and the native pronunciation of the contrast.

The current experiment shows that the effects of given factors on phonetic accommodation are highly dependent on the combination present during an interaction. In particular, it shows that controlling for the presence of exposure is highly important when studying the effects of feedback on phonetic accommodation. These results also confirm that the generalizability of computer-based approaches to real-life interaction is questionable.

3.5.5. Differences between human-human and human-computer settings

This study partly replicates previous research showing differences between humanhuman and human-computer phonetic accommodation using tasks without conversational goals (Burnham et al., 2010; Cohn et al., 2019; Raveh et al., 2019). Results in this study showed that the humanness of the interlocutor affects the degree of accommodation even when an interactive task is involved and also when the human confederate is played by a person sitting in the same room as the participant, instead of a virtual avatar (Staum Casasanto et al., 2010) or a video-presented confederate (Burnham et al., 2010). By using the same pre-recorded speech in the two settings we can discard that the differences found in the accommodation of speakers who interacted with a human or with a computer are linked to the acoustic properties of the voice. Instead, these differences can be linked to the presentation of the interlocutor as a human or as a computer and inevitably to the dimensions varying from one setting to the other. When presenting the interlocutor as a computer, participants immediately understand that the interlocutor's performance will be shaped by its limitations in adaptability. Participants would then only show accommodation when the communicative success was in danger, for instance when the computer showed signs of struggling, i.e., in the conditions where feedback was present. In the human-human settings, on the other hand, phonetic accommodation has been found to be more likely triggered by unmediated process; therefore, the differences in accommodation seemed to be triggered by other relevant factors, such as exposure. In addition, in the humanhuman setting, participants also had access to other non-verbal dimensions that are relevant in communication and which have an effect on the amount of phonetic accommodation displayed in human-human settings, such as eye contact (Schweitzer et al., 2017).

The findings of this study also have methodological implications. The current study follows other studies that move away from the use of non-interactive tasks in the study of phonetic accommodation (Cohn et al., 2019; Cohn & Zellou, 2019; Gessinger et al., 2017; Pardo et al., 2013; Rojczyk, 2013), by using tasks that allow for and encourage more interaction and spontaneity (Burnham et al., 2010; Gessinger et al., 2019; Staum Casasanto et al., 2010). Similarly to previous studies using more interactive tasks to compares these two settings, this study does not validate the generalizability of computer-based approaches to real-life interactions between human speakers, based on the differences found between human-computer and human-human settings. Hence, we also advocate for the implementation of more interactive human-human settings in the laboratory in order to study phonetic accommodation.

Perhaps surprisingly, we did not find evidence for exposure to a given sound in isolation to trigger phonetic accommodation in the human-computer setting. This finding contrasts with previous research using low demanding tasks, such as shadowing tasks, where exposure to a voice triggered phonetic accommodation to a computer interlocutor. A possible explanation could be that excluding conversational goals from the interaction encourages participants to plainly imitate the speech of the interlocutor. The puzzle-solving game used in this study includes an exchange of information between the interlocutors who have a common goal. In contrast, shadowing tasks, which have been previously used (Cohn et al., 2019; Cohn & Zellou, 2019; Pardo et al., 2013; Rojczyk, 2013), are arguably not representative of any type of interactive conversation. The conclusions of this study call for the implementation of more interactive settings in the study of phonetic accommodation in order to be bridge the gap between laboratory settings and real-life interactions. Making laboratory settings more similar to real-life interactions will make generalization of results possible.

This study highlights how important incorporating more fine-grained experimental control over exposure is when studying phonetic accommodation in laboratory settings. For instance, the comparison between the different conditions of this study, including feedback on the pronunciation of a vowel contrast, show that taking into account whether participants were exposed to the interlocutor's pronunciation of that contrast during the interaction is crucial to interpret the effect of feedback. In the literature dealing with exposure, this factor is often used as an umbrella term, including very broad interpretations, for instance, outside-of-the-lab uncontrolled exposure (Pardo et al., 2012), to very detailed accounts such as the recount of the number of tokens including the critical sounds that appeared during the conversation (Berry & Ernestus, 2018). In order to understand how exposure affects phonetic accommodation, researchers need to maintain the control over its presence in the experimental design. This study confirms not only that implementing this level of control in an interactive setting is possible but also that this level of control is indeednecessary in order to draw conclusions about the effect of other factors on phonetic accommodation.

The conclusions from this study also have some theoretical implications on the theory of phonetic accommodation. First, exposure in isolation did not trigger accommodations to the computer interlocutor in isolation and it also did not trigger accommodations when in combination with feedback in the human setting. This finding contrasts with literature claiming that phonetic accommodation could only be the result of mere exposure. If phonetic accommodation was the result of mere priming, being exposed to a cue should boost the level of activation of that cue and, thus, facilitate its use. However, Babel (2012), Hwang et al. (2015), Berry and Ernestus (2018), Gessinger et al. (2019) and the current study, among others, all show evidence for exposure triggering accommodation to only some, and not all, of the phonetic cues under study. Therefore, phonetic accommodation cannot be only the result of priming.

The effects of both factors under study here, both in isolation and in combination, were heavily dependent on the humanness of the interlocutor. Previous laboratory implementations have argued that stripping away the social component from the interaction can lead to a better understanding of phonetic accommodation (Pardo, 2013). However, the differences found in this study seem to show that including the human factor is essential when investigating phonetic accommodation and calls for an interpretation of phonetic accommodation as a socially-mediated linguistic process.

This first approach to study the differences in phonetic accommodation between human-human and human-computer settings in a controlled environment also has limitations that should be noted. Firstly, the generalizability of this study is, of course, restricted by its scope: in order to be able to study phonetic accommodation in a controlled way, decisions about the sounds included, the type of conversation, the gender of the participants and the interlocutor had to be made. Thus, when interpreting the conclusions of this study, we should take into account that the scope is restricted to those decisions made. In addition, none of the interlocutors in this study, human or computer, could show phonetic accommodation to the participants. Further research could determine whether manipulating the interlocutor’s recordings to show accommodation to the participants in other phonetic levels, such as speech rate or pitch, would affect the degree to which participants accommodate to the interlocutor.

Secondly, in order to maintain the experience of the interaction as homogeneous as possible to all participants, we decided to keep the feedback constantly negative in the twelve trials where participants had to produce the critical sounds. Therefore, participants were always corrected even if they tried to change their pronunciation of the contrast. The results obtained in this study, where feedback was shown to have an effect on the degree of phonetic accommodation, could differ if personalized feedback could be implemented in the interaction. For instance, using Automatic Speech Recognition in the set-up could enable the implementation of negative feedback only when required.

Finally, participants in this study were asked to rate the intelligibility and friendliness of the interlocutor. Surprisingly, we did not find any differences between the two settings or withing the different conditions. In addition, we found a very narrow range in the responses provided. These findings suggest that the scale used (1-5 scale) may have been too simple for capturing any differences between the settings or conditions. Further research studying the effect of the likeability of the interlocutor could incorporate more precise measurements in order to investigate whether phonetic accommodation is dependent on how friendly participants think the interlocutor is or how well they can understand them.

In conclusion, we have successfully compared phonetic accommodation in human-human and human-computer interactions using an interactive task and the exact same speech input in the two settings. The results clearly show that phonetic accommodation works differently when speakers interact with a human confederate than when they think they are interacting with a computer. Specifically, we have investigated how phonetic accommodation is affected by exposure and feedback in these two settings. We found that the two factors in isolation triggered phonetic accommodation in human-human settings, whereas the combination of the two factors was more effective fostering accommodation to the computer interlocutor. These results call into question the generalizability of conclusions on phonetic accommodation using computer-based approaches.

Chapter 4

L2 learning through phonetic accommodation:

the effect of exposure and feedback on intelligibility and nativelikeness

ABSTRACT

This study investigates L2 learning in conversation through accommodation. We examine whether Dutch learners of English adapted their pronunciation of the English contrast /æ, ɛ/, which they tend to merge, after receiving (1) exposure to the native-like pronunciation of these vowels and/or (2) feedback on their pronunciation of the contrast during an interaction. Perceptual measurements on the speakers pre- and post-interaction productions of the vowels were collected in three experiments. Experiment 1 investigated whether native listeners perceived an improvement in the intelligibility of the learners' pronunciation after the interaction. Experiment 2 examined whether the same listeners perceived an increase in the nativelikeness. Experiment 3 explored whether Automatic Speech Alignment (ASA) technology detected an improvement in the intelligibility of the learners' pronunciation. Results show that exposure and feedback in isolation and, even more so, in combination fostered a more intelligible pronunciation. Learners sounded overall more native-like after the interaction, but no additional increase was perceived as the result of exposure and feedback. In fact, the combination of the two factors hindered this overall improvement in nativelikeness. Interestingly, the ASA-based intelligibility was not affected by exposure or feedback at all. In conclusion, this study highlights the importance of using different perceptual measurements to assess L2 learning as the result of phonetic accommodation, and calls for more fine-tuning of ASA technology in order to replace human listeners' judgements.

4.1. Introduction

Dialogue is an essential context in the acquisition of a native language (L1), but also in the process of learning a second language (L2). In fact, learning through interaction was reported to be the second most popular method for L2 acquisition in most European countries, only after formal instruction (Eurobarometer, 2012). Furthermore, engaging in conversations in the L2 is not only the means, but often the main goal of learning a new language. However, research on L2 learning typically focuses on other contexts of L2 learning; where, for instance, speakers learn the L2 in a classroom setting (Loewen & Philp, 2006; Lyster, 1998; Mackey et al., 2007; Mackey & Gass, 2006, among others) or use software designed for the same purpose (Cucchiarini et al., 2012; Nushi et al., 2017; Saito & Lyster, 2012, among others). Little attention is given to the learning that takes place when speakers use the L2 in interactions.

L2 learning through interaction could be explained as the result of accommodation (Giles & Ogay, 2007), also known as alignment (Pickering & Garrod, 2004) or entrainment (Levitan et al., 2012). This phenomenon refers to the fact that during interactions, speakers can change the characteristics of their speech to adopt those displayed in the speech of the interlocutor (Babel & Bulatov, 2012; Berry & Ernestus, 2018; Cohn et al., 2019; Lewandowski & Nygaard, 2018; Pardo, 2006; Pardo et al., 2010). Accommodation can affect different linguistic levels, for instance, speakers may adapt their vocabulary, syntax or pronunciation to match the interlocutor's speech. This chapter focuses on how phonetic accommodation to specific sounds during an interaction can result in the learning of the L2 pronunciation.

When speakers interact in their L2, they can make adaptations to their pronunciation of sounds in their attempt to align to the pronunciation of the interlocutor, who could be, for instance, a native speaker of the L2. These adjustments to their pronunciation could survive after the interaction and, therefore, result in L2 sound learning. Originally accommodation in non-native speakers was theorized to be more difficult and less likely to take place than accommodation in situations involving exclusively native speakers (Costa et al., 2008). However, recent literature has provided overwhelming evidence showing that phonetic accommodation can and does take place in interactions including non-native speakers (Berry & Ernestus, 2018; Hwang et al., 2015; Lewandowski & Nygaard, 2018; Rojczyk, 2013; Simonet, 2014). This chapter aims to contribute to understand phonetic accommodation in L2 context by identifying which factors trigger this phenomenon.

4.1.1. The mechanisms underlying phonetic accommodation

In the quest of identifying the mechanisms underlying phonetic accommodation, the effects of various different factors on this phenomenon have been studied (e.g., Auer et al., 1998; Babel & Bulatov, 2012; Gessinger et al., 2019; MacLeod, 2014; Pardo, 2006; Pardo et al., 2010; Chapter 3). For instance, the degree of phonetic accommodation may be affected by the speaker's gender (Namy et al., 2002; Willemyns et al., 1997), the perceptual salience of the sound (Auer et al., 1998; MacLeod, 2014), degree of exposure (Cohn et al., 2019; Wanrooij et al., 2013, Chapter 3) and presence versus absence of feedback (Burnham et al., 2010; Chapter 3). The current study investigates the effect of controlled exposure and feedback on phonetic accommodation in L2-L1 interactions, focusing on whether the interlocutor can perceive the adaptations a speaker may implement in their speech as the result of receiving feedback and/or exposure.

4.1.1.1. The effect of Exposure on phonetic accommodation

Accommodation, understood as alignment, has been proposed to be the result of priming (Pickering and Garrod, 2004). According to this explanation, when speakers receive exposure to a given linguistic cue in interaction (for example, exposure to a specific word), the boosted level of activation of that cue would increase the chances of the speakers using it in the near future. Literature investigating this account has reported that mere exposure can indeed foster phonetic accommodation to the interlocutor's pronunciation. For instance, speakers have been found to adapt their speech after mere exposure to match the interlocutor's speech rate and pitch level (Eijk et al., 2019; Gijssels et al., 2016; Levitan & Hirschberg, 2011; Staum Casasanto et al., 2010), and also the intonation contours used by the interlocutor (Gessinger et al., 2019). Similarly, speakers have also been found to adapt their pronunciation of specific segments when they receive exposure to the interlocutor's pronunciation of those sounds, including consonants (e.g., Gessinger et al., 2019) and, perhaps more vastly documented in the literature, vowels (e.g., Babel, 2010, 2012; Babel et al., 2013; Berry & Ernestus, 2018; Pardo et al., 2012; Chapter 3).

Previous literature studying the effect of exposure on phonetic accommodation tend to make use of very artificial settings to control for the exposure participants receive (e.g., Eijk et al., 2019; Gessinger et al., 2019; Gijssels et al., 2016; Levitan & Hirschberg, 2011, Gessinger et al., 2019; Staum Casasanto et al., 2010). The use of this type of settings hinders the generalization of the findings about exposure to other, more natural and spontaneous contexts. When studying the effect of exposure on phonetic accommodation in spontaneous settings with a human interlocutor speaking live, on the other hand, no control can be exercised over phonetic input participants receive (e.g., Berry & Ernestus, 2018). Literature researching accommodation affecting other linguistic levels have managed to combine both spontaneity and control over the input. For instance, the Scripted Confederate methodology (Branigan et al., 2000), where a confederate acts out a script without participants noticing, has been successfully used to study syntactic accommodation in a spontaneous setting while controlling for the syntactic structures participants are exposed to (Kootstra et al., 2010). However, in the phonetic level, where subtle phonetic differences are relevant, implementing this level of control in a spontaneous setting has not been possible so far.

4.1.1.2. The effect of Feedback on phonetic accommodation

From a listener-driven perspective, accommodation can be argued to be the result of speakers trying to ensure understanding in communication (Clark, 1996). For instance, by using the same set of speech properties as the interlocutor uses, a speaker can guarantee that the interlocutor will better understand the message they are trying to convey. In order to investigate this perspective, research often implements feedback implying lack of understanding on behalf of the interlocutor (Burnham et al., 2010; Chapter 3). When a speaker receives feedback from the interlocutor suggesting that they misunderstood a part of the message due to differences in pronunciation, the speaker may adapt their pronunciation as a result. This context becomes even more relevant in situations involving a non-native speaker of a language. Due to differences between a learner's L1 and the L2 sound systems, there can be mismatches in the pronunciation of specific sounds. These mismatches can sometimes lead to disruptions in the communication (Flege, 1993).

The effect of feedback on L2 learning is, most typically studied in classroom and L2-training software settings (Adams et al., 2011; Cucchiarini et al., 2012; Dlaska & Krekeler, 2013; Lee & Lyster, 2017; Nassaji, 2009; Saito & Lyster, 2012). Feedback has been found to be overall effective in classroom and computer-based settings. For example, learners of German as a second language have been found to show more learning in L2 pronunciation when receiving individual corrective feedback on top of performing listening activities in German (Dlaska & Krekeler, 2013). The effect of feedback in live interactions, however, is currently understudied (e.g., Burnham et al., 2010; Chapter 3).

4.1.1.3. The effect of Feedback and Exposure in combination

Combining feedback from a human interlocutor in a spontaneous conversation while also controlling for the phonetic exposure participants receive can pose a challenge. As a result, research investigating the effect of feedback on phonetic accommodation rarely disentangles the influence of feedback from the influence of exposure to the interlocutor's pronunciation on phonetic accommodation (for example, Burnham et al., 2010, Chapter 3). Thus, it is unclear whether the effect of negative feedback on the L2 speakers' pronunciation differs depending on whether the L2 speakers are provided with evidence of the target-like pronunciation of the sounds on which they have received feedback.

4.1.2. Studying the effect of exposure and feedback on phonetic accommodation maintaining the experimental control

Studying phonetic accommodation typically comes with either the sacrifice of the control over the input participants receive or the detriment of the spontaneity of the interaction. When the spontaneity of the conversation is prioritized over the control, phonetic accommodation can be studied in contexts where two participants freely interact with each other (for instance, Berry & Ernestus, 2018). As a consequence, the researcher has no control over the amount or the quality of the input participants receive. Maintaining the control over the input is possible by means of using pre-recorded speech. However, the implementation of pre-recorded audio typically comes with the sacrifice of the spontaneity, the simplification of the task (for instance, Cohn et al., 2019; Pardo, 2006; Pardo et al., 2012), and, oftentimes, the dehumanization of the confederate interacting with the participant. The confederate, then, is presented as a voice (Cohn et al., 2019) or virtual reality avatar (Staum Casasanto et al., 2010). These sacrifices crucially may affect the conclusions drawn from these studies because the presentation of the confederate, as a human-like or a computer-like entity, may trigger clear differences in the degree of phonetic accommodation participants display (Branigan et al., 2010; Cohn et al., 2019; Chapter 3).

Here, we aim to fill in the gaps in the literature investigating exposure and feedback, in isolation and in combination, in a context combining both phonetic control over the input participants receive, and the spontaneity of the interaction, by using the Ventriloquist Paradigm (Chapter 2). This paradigm allows for the controlled presence or the absolute absence of exposure to specific sounds during the interaction. Therefore, the effect of exposure to the interlocutor's pronunciation can be assessed by comparing interactions where L2 speakers receive either no exposure at all or controlled exposure to native-like pronunciations of the critical sounds. In addition, this method, where pre-recorded speech is used, ensures that all participants who receive exposure are exposed to the same phonetic input.

This level of fine-grained control over the phonetic input also allows to disentangle the effect of feedback from exposure. In particular, by controlling the presence and absence of exposure using pre-recorded speech, the effect of feedback on phonetic accommodation can be assessed when L2 learners receive exposure to the native-like pronunciation of the sounds and when they do not receive evidence of how native speakers produce the sounds on which they are receiving feedback.

4.1.3. Using perceptual measurements to assess phonetic accommodation

Typically, in order to assess phonetic accommodation, recordings of the participants' speech are analyzed to determine whether speakers' pronunciation changed from before to after the interaction, or even during the interaction. Changes in the pronunciation are usually assessed by either acoustic analyses or perceptual measurements (Pardo et al., 2013).

Acoustic analyses, such as the ones used in Chapter 3, involve the measurement of a given phonetic cue present in the audio recordings of participants' speech. Using acoustic analyses offers a very fine-grained understanding of the adaptations speakers make to specific phonetic cues in isolation, for instance speech rate, pitch or the height and frontness of vowels. However, the study of phonetic cues to determine whether phonetic accommodation takes place does not provide information about whether the interlocutor could perceive a change (Pardo et al., 2013). In addition, the analysis of different phonetic cues is usually done in isolation, which makes it difficult to determine how the cues interact in the perception by the interlocutor.

Perceptual measurements typically involve new naïve listeners judging the audio recordings of the speakers participating in the interaction. For instance, listeners may be asked to determine whether a speaker's pronunciation is more similar to the interlocutor before or after the interaction (Pardo et al., 2013). Traditionally, AXB similarity tasks are used in order to assess whether speakers aligned their pronunciation to become more similar to the interlocutor. Thus, perceptual measurements can help directly determine whether the adaptations made during the interaction could have been perceived by the interlocutor as phonetic accommodation. Furthermore, listeners judging the audio recordings tend to use all the phonetic cues available for their decisions; thus, perceptual judgments are not typically restricted to a single phonetic cue but to the overall combination of all cues available.

4.1.3.1. Automatically-generated intelligibility

Automatic Speech Recognition (ASR) can be used to evaluate changes in pronunciation in a more efficient and objective way than collecting human judgements (Cucchiarini et al., 2012; Escudero-Mancebo et al., 2015; Franco et al., 2010; Van Nuffelen et al., 2009). For instance, in the field of dysarthria, Automatic Speech Alignment (ASA) systems have been proposed in order to objectively detect deviant pronunciation in patients as a substitute for the perceptual judgement of clinicians (Van Nuffelen et al., 2009). In the field of L2 acquisition, automatic methods have also been implemented in order to objectively measure the process of learning L2 sounds (e.g., Cucchiarini et al., 2012; Franco et al., 2010). The development of the EduSpeak toolkit (Franco et al., 2010) allows for the use of speech recognition technology for pronunciation scoring purposes. The advantage of using this technology is that learners can receive real-time, online, and completely personalized feedback on their pronunciation of the L2 sounds. Furthermore, ASA technology makes the process of evaluating L2 learning substantially more efficient and objective than collecting the judgement of human listeners.

However, the validation of ASA/ASR-based technology to replace human judgements is yet to be determined. The justification of the use of automatically-generated judgements tends to be based on the comparison between the artificial judgements and judgements provided by very limited samples of human listeners. For example, the

EduSpeak toolkit was evaluated by comparing the performance of the software to four transcribers (Franco et al., 2010). Moreover, these small samples of human listeners tend to be trained phoneticians, who may not be representative of the general population of speakers.

4.1.4. This study

The current study is based on a data set acquired in Chapter 3, targeting the English vowel contrast /æ, ɛ/ produced by Dutch speakers before and after they interacted with a native speaker of English. This vowel contrast has been reported to be difficult for Dutch learners of English, who tend to merge the two vowels into an in-between sound (Elsendoorn, 1985; Wang & van Heuven, 2006). The current study aims to investigate whether exposure and feedback, in isolation and in combination, foster L2 learning in the context of an interaction from a perceptual perspective. In order to do so, perceptual measurements are used to assess whether Dutch learners of English aligned to the native pronunciation of the English /æ, ɛ/ contrast, after participating in an interaction with a confederate. The speakers in the dataset did or did not receive exposure to the native pronunciation of the vowel contrast and/or feedback on their pronunciation of these sounds during an interaction. Thus, the presence and absence of exposure and feedback were fully crossed, resulting in four different conditions: a Control condition, where participants did not receive exposure or feedback; an Only Exposure condition, where participants received exposure but no feedback; an Only Feedback condition, where participants received feedback but no exposure; and a Combined condition, where participants received both feedback and exposure.

The dataset was collected using the Ventriloquist paradigm in order to manipulate the presence and absence of exposure. This paradigm allows for the implementation of pre-recorded speech in spontaneous interactions, unbeknownst to participants, who think they are having a live conversation with a confederate. The presence of exposure meant the inclusion of 12 tokens of each of the critical vowels, /æ, ɛ/, in the interaction in the context of minimal pairs, for instance, “bad”-“bed”. The absence of exposure was implemented by carefully scripting the pre-recorded speech of the confederate avoiding any token of the critical vowels. The presence of feedback was implemented as visual feedback on the screen making participants aware that the confederate had misunderstood their pronunciation of a word including /æ/, e.g., “bad”, for a minimal pair word including the /ɛ/ vowel, e.g., “bed”. In the conditions without feedback, participants were never made aware of any miscommunications triggered by their pronunciation of the critical contrast. The L2 speakers participating in this experiment performed identical production tasks before and after the interaction serving as the baseline and the post-interaction pronunciations of the English vowel contrast, /æ, ɛ/.

In Chapter 3 this data set was analyzed from an acoustic point of view. Phonetic accommodation was measured as any change in the first (F1) and second (F2) formants of the two critical vowels as indication of the Dutch speakers modifying the height and frontness, respectively, of these two sounds. Results of these acoustic analyses show that exposure to the interlocutor’s pronunciation of the critical vowels triggered Dutch learners to adapt their production of the /æ, ɛ/ contrast: in the post-test participants enlarged the difference in frontness between the two vowels. Similarly, the feedback in isolation, i.e., without exposure to the native-like pronunciation of the contrast, also triggered adaptations in the learners’ pronunciation. In this condition, the non-native speakers adapted the pronunciation of the two vowels both in terms of height and frontness. The combination of the two factors together did not seem to affect the learners’ pronunciation of these two vowels in terms of height and frontness. As Chapter 3 included only an acoustic study, it is unclear whether the changes detected can be perceived by the human ear.

The current study aims to assess phonetic accommodation from a perceptual point of view in order to determine whether the adaptations triggered on L2 speakers’ pronunciation by exposure, feedback, and the combination of the two factors together during an interaction could be perceived by the interlocutor. Thus, native speakers of English were asked to judge the recordings of Dutch learners of English producing the critical contrast /æ, ɛ/ in order to confirm whether they aligned to the native speaker. Studying situations involving non-native speakers of an L2 expands the range of perceptual tests that can be used. More than L1 speakers, L2 learners can show accommodation to the native speaker's speech by improving their intelligibility or nativelikeness of their pronunciation. This study uses intelligibility and nativelikeness judgements to determine whether the presence of exposure, feedback in isolation, and in combination triggered L2 speakers to adopt a more intelligible and native-like pronunciation of the L2.

In addition, the current study aims to contribute to the literature assessing the use of automatically-generated intelligibility scores to measure L2 learning. In order to do so, an ASA-based system is used to assess the degree of phonetic accommodation in L2 speakers after an interaction with a native speaker of the L2. The degree of phonetic accommodation is measured as any improvement in the intelligibility of the pronunciation of the difficult L2 contrast, /æ, ɛ/. The use of the automatically-generated scores is analyzed by first determining the degree of agreement between the ASA-system and a large sample of human native speakers of English when evaluating the intelligibility of the L2 speakers' pronunciation. Second, we analyzed whether ASA-system would detect the same adaptations in L2 learners' pronunciations, resulting from exposure and feedback, as native speakers of the L2.

In summary, this study investigates whether native speakers and an ASA-system can detect adaptations in pronunciation that L2 learners apply after receiving exposure to the native-like pronunciation of a contrast and feedback on their own pronunciation of the same contrast. Thus, the three research questions of this study are:

> RQ1: Is L2 speakers' pronunciation perceived as more intelligible by native listeners as the result of the L2 speakers receiving exposure and/or feedback during an interaction?

RQ2: Is L2 speakers' pronunciation perceived as more nativelike by native listeners as the result of the L2 speakers receiving exposure and/or feedback during an interaction?

RQ3: Is L2 speakers' pronunciation perceived as more intelligible by ASA-based technology as the result of the L2 speakers receiving exposure and/or feedback during an interaction?

Each question was addressed in an experiment. In Experiment 1, native speakers of English judged the intelligibility of L2 learners' pronunciation of the /æ, ɛ/ vowels, extracted from minimal pair words participants read in a production task, before and after the L2 learners had received exposure and/or feedback during an interaction. The intelligibility was based on whether the native speakers could correctly identify the vowel intended by the L2 speakers. In Experiment 2, the same native speakers provided nativelikeness judgements of the same pre- and post-interaction /æ, ɛ/ tokens they also evaluated in Experiment 1. The native speakers were asked to use a 5-point scale, ranging from native-like to strongly accented. In Experiment 3, ASA-based technology was used in order to generate artificial intelligibility scores with a procedure similar to the one used to collect intelligibility from the group of native speakers in Experiment 1. A forced aligner using English pre-trained acoustic models of phones was used to identify the vowels produced by the Dutch speakers before and after the interaction.

Previous research have reported that the presence of exposure triggers phonetic accommodation from the speaker to the interlocutor's pronunciation (Cohn et al., 2019; Hwang et al., 2015, among others). The acoustic analysis revealed that the L2 speakers of this study accommodated their pronunciation of the /æ, ɛ/ contrast to the native speaker (Chapter 3). Thus, in Experiment 1, the native speakers' accuracy at identifying the vowel produced by the Dutch learners was hypothesized to be higher in the post-test than in the pre-test, after learners had received exposure during the interaction. Similarly, the presence of feedback during an interaction has also been found to trigger adaptations in the speakers' pronunciation (e.g., Burnham et al., 2010; Chapter 3). Therefore, the intelligibility was also hypothesized to increase from the pre-test to the post-test task as the result of the L2 learners receiving feedback on their pronunciation of the critical contrast. The combination of the two factors when studied in controlled interactions, however, has been found to not trigger phonetic accommodation, and specifically in the dataset under study (Chapter 3). Thus, the intelligibility was not hypothesized to change from the pre-test to the post-test after L2 speakers had received both exposure and feedback during the interaction.

Similarly to the hypotheses for Experiment 1, given that previous literature has found that both exposure and feedback, in isolation, trigger adaptations in the pronunciation of L2 learners, the native-likeness ratings on the L2 speakers' pronunciation of the critical contrast, /æ, ɛ/, in Experiment 2 were hypothesized to increase from pre-test to post-test when the participants had received exposure to the native-like pronunciation and corrective feedback on their pronunciation. However, comparably to the hypotheses for Experiment 1, the nativelikeness ratings were not hypothesized to change from pre-test to post-test when the two factors were combined during the interaction (Chapter 3).

For Experiment 3, the intelligibility scores by the ASA-system were predicted to behave identically to those collected from human native speakers in Experiment 1. The agreement between the ASA-system and the native speakers were expected to be similar to the agreement among human listeners. The ASA-based technology was therefore also hypothesized to detect an increase in the intelligibility of the critical contrast, after Dutch speakers had received either exposure to the native-like pronunciation of the critical contrast or feedback on their pronunciation of the contrast. The combination of the two factors together during the interaction was hypothesized to not affect the degree of phonetic accommodation as judged by ASA-based technology.

For experiments 1 and 2, native speakers' intelligibility and the nativelikeness judgements on the L2 speakers' pronunciation were collected in a single experimental session. Participants first performed the intelligibility judgement task and, then, the nativelikeness judgement task. Between the two tasks, participants could take a short break of approximately 5 minutes. For Experiment 3, a forced aligner, Montreal Forced Aligner (MFA) (McAuliffe et al., 2017), was used in order to generate intelligibility scores for L2 speakers' pronunciation of the critical contrast.

4.2. Speech material

The three perceptual experiments included in this study used recorded speech materials from a production experiment where L2 learners interacted with a confederate, collected in Chapter 3. This section briefly explains the characteristics of the L2 speakers, the materials and the procedure of the experiment, and the characteristics of the audio tokens selected for the three perceptual experiments included in the current study.

4.2.1. Speaker participants

Forty-eight native speakers of Dutch (M_{age} = 21.4 years, sd = 2.3), who had been raised monolingually, participated in this production experiment. They were all female students at Radboud University Nijmegen (Netherlands). Their average score on the LexTale test (Lemhöfer & Broersma, 2012) was 75.78 (sd = 10.82), which is equivalent to an upper intermediate (B2) level in the Common European Framework of Reference for Languages (Council of Europe, 2001). None of the participants reported any speech or learning impairments and all of them were rewarded for their time with either study credits or a voucher.

4.2.2. Materials and procedure

4.2.2.1. Implementation of the Ventriloquist Paradigm

The 48 L2 speakers participated in an interaction with a human confederate. The interaction was based on a puzzle-solving task that required the participant and the confederate to share information on their screens, and collaborate. We developed a new paradigm, the

Ventriloquist paradigm (Chapter 2) to covertly implement pre-recorded speech in an interactive setting unbeknownst to participants. Thus, the contributions the confederate made to the conversation were not live utterances but had been scripted, pre-recorded and were played to the participants during the interaction. In order to maintain the illusion that the speech was produced live, both the confederate and the participant were equipped with a microphone and a pair of headphones. The participants were told that they would hear the confederate's live speech through their headphones but, in reality, they listened to the pre-recorded samples of speech.

During the interaction, the confederate was in control of playing the pre-recorded speech samples, which she triggered by pressing buttons in a numeric keyboard, hidden from the participant's sight. She could choose from a diverse range of different utterance categories to play. Each button in the pad was connected to a different category of speech and included approximately 30 different audio recordings conveying the same meaning with different phrasings and intonations. The categories included speech utterances related to the puzzles presented on the screen in each trial and "spontaneous" speech categories used to maintain the illusion of a spontaneous live conversation. The spontaneous speech categories consisted of affirmative responses (e.g., "yup", "uh-huh"), negative responses (e.g., "no", "I don't think so"), requests for repetition, among others (see Chapter 2 for a complete description).

4.2.2.2. Implementation of Exposure and Feedback

During this interaction, the presence and absence of exposure to the native-like pronunciation of the critical contrast, /æ, ɛ/, and feedback on the learners' pronunciation of the contrast were fully crossed, resulting in four different conditions: the Control Condition, the OnlyExposure Condition, OnlyFeedback Condition, and Combined Condition. See Table 1 for an overview.

Table 1. Overview of the four conditions of the experiment used to elicit the speech material

	No exposure	Exposure
No feedback	Control Condition	OnlyExposure Condition
Feedback	OnlyFeedback Condition	Combined Condition

The implementation of these two factors in the interaction was based on two different types of puzzle-solving trials: the Semantic Relation trials, designed to implement controlled exposure; and the Code Breaker trials, designed to implement feedback.

Implementation of Exposure in interaction

In order to implement controlled exposure to the critical contrast in the interaction, the pre-recoded speech of the confederate had been carefully designed to avoid the presence of any word including /æ, ɛ/ in both the puzzle-related and the spontaneous speech audio samples that the confederate could play. The only context in which these two vowels could appear was the Semantic Relation trials. A total of 12 Semantic Relation trials were included during the interaction. In each of these trials, the participant saw a picture on the screen, for instance, a picture of a landscape with hills (see Figure 1). The confederate saw four words on her screen, consisting of 2 sets of minimal pair words, e.g., "bad", "bed", "hill" and "heal" (see Figure 1). Only one of the four words displayed on the confederate's screen was semantically related to the picture on the participant's screen, in this example, "hill". Thus, the goal of these Sematic Relation trials was for the two players to share information on their screens in order to discover which of the words on the confederate's screen was related to the picture on the participant's screen.

Figure 1. Visual representation of the Semantic Relation trials used to implement exposure, including the confederate's screen (left) and the participant's screen (right).

Each Semantic Relation trial started with the participant describing the picture displayed on their screen. The confederate, then, played a pre-recorded audio sample where she read out loud the four words displayed on her screen and decided which one was semantically related to the description provided by the participant (e.g., "the words in my screen are bad, bed, hill, heal, so the right one must be hill").

In the conditions with exposure, the 12 Semantic trials were considered to be critical. Thus, in the 12 trials the words that the confederate read out loud included an /æ, ɛ/ minimal pair (e.g., "The words on my screen are **bad, bed**, hill and heal, so the right one must be hill"). In order to ensure participants were listening to the critical words, in these trials the /æ, ɛ/ words were always presented as the first two words read out loud. In order to ensure that the exposure was passive, the critical words were never related to the picture; thus, participants did not need to make any decisions about them. The order of presentation of the /æ, ɛ/ words, i.e., whether they were presented as, for instance, "bad", "bed" or as "bed", "bad", was counterbalanced throughout the experiment. In conditions without exposure, the /æ, ɛ/ minimal pair words, e.g., "bad"-"bed", were replaced by filler words, e.g., *bin-pin*. That is, the audios the confederate played in Semantic Relation trials never contained the critical contrast (e.g., "The words in my screen are pin, bin, hill and heal, so the right one must be hill"), and participants were never exposed to the /æ, ɛ/ contrast during the interaction.

Implementation of Feedback in interaction

The implementation of the presence versus absence of feedback was based on the Code Breaker trials. During these trials, the participant saw four different shapes on their screen, each shape was linked to a word displayed precisely below the shape. The confederate could see a sequence of several shapes followed by a question mark at the top half of the screen and a set of four words in the half below (see Figure 2). These four words consisted of two sets of minimal pair words and they were the same as the ones displayed on the screen of the participant.

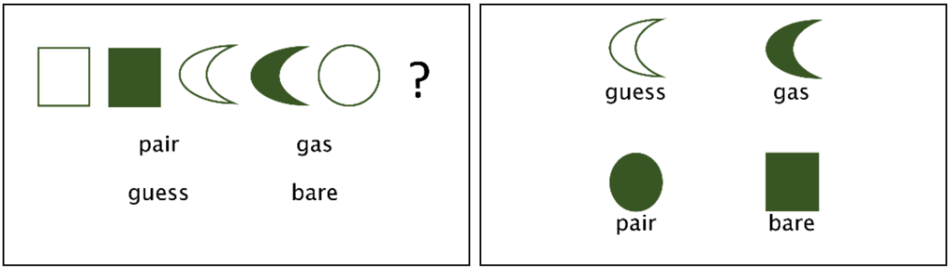


Figure 2. Visual representation of the CodeBreaker trials used to implement feedback, including the confederate's screen (left) and the participant's screen (right).

The confederate and the participant collaborated in order to find which shape on the participant's screen was linked to the sequence of shapes in the confederate's screen. Every trial started with the confederate playing an audio sample with a description of the sequence of shapes, e.g., "I think we need a green circle…."Once the participant and the confederate agreed on which of the participant's shapes matched the incomplete sequence, the participant instructed the confederate to click on the word that was linked to the shape completing the sequence to proceed to the next trial. Out of the total 60 trials, only 12 of them were considered critical trials. In these trials, the word linked to the right shape on the participant's screen contained the /æ/ vowel and it was displayed on the screen in the context of an /æ, ɛ/ minimal pair, for instance, "bad"-"bed". Thus, participants had to pronounce this word to instruct the confederate to click on it, and their pronunciation needed to be distinct enough for the confederate to be able to distinguish from the word on the screen with /ɛ/.

In conditions with feedback, the participant could see a grey rectangle displayed for 3.0 seconds around the word selected by the confederate on their screen. This way, the participant was able to see which word the confederate had clicked on. The confederate always clicked on the word pronounced by the participant except in the critical trials, i.e., where the target word contained /æ/. In these critical trials, the confederate deliberately clicked on the /ɛ/ minimal pair word option also displayed on the screen, e.g., “bed”, implying that the participant’s pronunciation of /æ/ had been perceived as /ɛ/. This visual feedback was maintained homogeneously negative throughout the experiment, in all 12 critical trials, regardless of the performance of the participants.

In conditions without feedback, after the confederate clicked on one of the words, the next trial started automatically. The grey rectangle did not appear on the screen, so the participant could never see the selection of the confederate and, consequently, they were never made aware of whether the confederate had misinterpreted their pronunciation of /æ/ for /ɛ/.

4.2.2.3. Assessing phonetic accommodation of the pronunciation of /æ, ɛ/

Participants performed identical production tasks immediately before and after the interaction. Both the pre-interaction and the post-interaction tasks consisted of a sentence completion task. Participants were presented with incomplete sentences, one by one, and they were instructed to, first, think of a suitable ending for the sentence and then read the sentence together with their proposed ending out loud.

A set of 24 /æ, ɛ/ minimal pair words were included in the sentences participants had to complete in both the pre-interaction and post-interaction tasks. These items were high frequency words (M = 4.81, sd = .661), according to SUBTLEX-UK (Van Heuvel et al. 2014) (see Appendix 1 for the list of words). Each word was embedded in two matched sentences, which were designed to be as similar as possible in terms of meaning, syntactic structure and the intonation expected, for instance, (1) “I have some **bad** news for you, you… and (2) You have some **bad** habits, such as….” One of the matched sentences was used for the pre-test task and the other one was used for the post-test task in order to ensure the context in which the same vowel had been produced before and after the interaction was comparable. Which of the matched sentences appeared in the pre-test or post-test was counterbalanced across participants. Every task consisted of a total of 48 sentences. The order of presentation of the sentences was randomized. The sentences overall varied in difficulty to complete and also in length.

4.2.2.4. Word selection for the perception experiments

A total of 1,101 audio recordings of words were extracted from the sentences the L2 learners produced in the Sentence Completion production tasks described above. The words selected consisted of the 24 /æ, ɛ/ minimal pair words embedded in the sentences participants had to complete. Every participant from the production study described above was represented in the selection of tokens by at least 24 tokens, which consisted of 6 pre-test /æ/ tokens, 6 pre-test /ɛ/ tokens, 6 post-test /æ/ tokens and 6 post-test /ɛ/ tokens. The same set of 1,101 extracted word tokens were used in all three perceptual experiments.

4.3. Experiment 1: Intelligibility by human listeners

The aim of Experiment 1 is to determine whether native listeners can perceive an improvement in the intelligibility of L2 speakers’ pronunciation of the critical contrast, as the result of the L2 speakers receiving exposure and/or feedback during the interaction.

4.3.1. Methods

4.3.1.1. Participants

A total 107 native listeners of American English, who had been raised monolingually, (79 females, $M_{age} = 19.24$, $sd_{age} = 1.19$) were recruited as listeners for this study at the University of Kansas (Lawrence, Kansas, U.S.A.). None of the participants reported any hearing, reading or learning impairments. All of them were rewarded for their time with study credits.

4.3.1.2. Materials

The 1,101 word tokens extracted from the production task were divided into three different lists. Each list included a total of 387 tokens. Every token occurred in only one list, except for a subset of 13 tokens that were present in all three lists. These 13 tokens were randomly selected and the goal of this subset was to enable a direct comparison of the agreement among the listeners across lists. Listeners were randomly assigned to one of the three lists (list 1: 36 listeners; list 2: 36 listeners; list 3: 35 listeners). The tokens included in each list were divided into 48 blocks, every block consisted of 8 tokens produced by the same speaker. Of the eight tokens, four had been extracted from the pre-interaction task and the other four from the post-interaction task. In every block, four tokens contained the /æ/ vowel and the other four the /ɛ/ vowel. Although the word tokens were repeated throughout the experiment, the words included in each block were different from each other in order to avoid listeners attempting to be consistent with their immediate previous judgements of the same token.

4.3.2. Procedure

The Intelligibility task was programmed using ROLEG, an application designed to present visual and auditory stimuli to participants developed by Radboud University Nijmegen. In order to perform the task, participants were equipped with headphones and a keyboard. At the beginning of every trial, an audio recording of a word was automatically played to the participants through the headphones. At the same time, participants were visually exposed to two possible orthographic transcriptions of the word played. The task of the participants was to identify which of the two orthographic transcriptions corresponded to the word they had heard. The orthographic transcriptions displayed on the screen were the correct orthographic transcription of the word intended by the Dutch learner, for instance, “bad”; and the minimal pair word containing the other vowel of the critical /æ, ɛ/ contrast, “bed”. On the screen of each trial, participants also saw which key on the keyboard was linked to each orthographic transcription option. To provide their answers, participants pressed the key linked to the word they thought they had heard. Immediately after pressing the key, a new trial started. After every block consisting of eight trials, a screen instructed participants to take a self-paced small break. The whole experiment took approximately 15 minutes.

4.3.3. Results and discussion

The responses given by the native listeners of English in the intelligibility task were coded into a binary variable, reflecting whether the native listeners identified the word intended by the Dutch speaker as such, e.g., “bad”, or whether they indicated they thought they heard the counterpart including the other critical vowel, “bed”.

4.3.3.1. Reliability assessment

In order to assess the reliability of agreement in the responses by all the listeners who participated in this study, the Fleiss’ kappa was calculated using the *kappam.fleiss* function from the “irr” package (Gamer et al., 2019) (version 0.84.1) in R (version 3.5.0)(R Development Core Team, 2008). First, the agreement between all the participants who performed this task was assessed. In order to do so, a Fleiss’ kappa was calculated for all listeners using their responses to the 13 word tokens that were present in the three lists used and, therefore, had been judged by all 107 listeners. The kappa of this comparison shows slight agreement between the listeners (kappa = 0.19).

In order to determine whether the low agreement was related to the different lists to which participants had been assigned, the degree of agreement among listeners who were assigned to the same list was also calculated, also using the Fleiss' kappa, for each list separately taking into account the 387 tokens included in each different list. The results of the kappa's calculated for the listeners assigned to the same list also revealed slight agreement between the raters for the first (kappa = 0.29), second (kappa = 0.33) and third list (kappa = 0.30).

4.3.3.2. Detecting phonetic adaptations in terms of intelligibility as the result of receiving exposure and feedback

Figure 3 shows the percentage of tokens that the native listeners correctly identified as containing the vowel intended by the Dutch learners. The accuracy is grouped by test (pre-interaction and post-interaction), vowel (/æ/ and /ɛ/) and condition (Control, Only Exposure, Only Feedback and Combined). Overall, the average of correctly identified tokens seems to be higher for the /ɛ/ word tokens than for the /æ/ word tokens. This difference suggests the Dutch-accented pronunciation of /æ/ is less intelligible than /ɛ/ for native listeners of English.

For the /æ/ tokens, the accuracy seems to be substantially higher in the post-test than in the pre-test for the Only Exposure, Only Feedback and the Combined condition. Especially in the Combined condition, this pre-test post-test difference in accuracy seems to be particularly large. In the Control condition, in contrast, this increase in accuracy from the pre-test to the post-test seems to be less evident. As regards /ɛ/, the average percentage of tokens identified as the intended vowel displays only minor differences between the pre-test and the post-test responses. This lack of improvement seems to be present in all of four conditions.

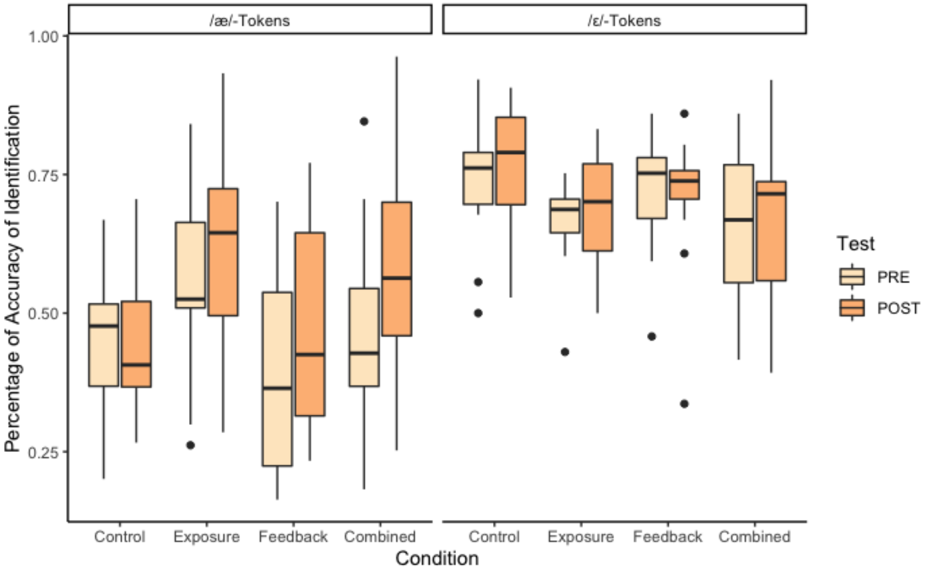


Figure 3. Boxplots representing the distribution of the speakers' average percentage of tokens correctly identified by the listeners, before (light orange) and after (dark orange) the interaction, divided per condition and vowel. In this figure, and also in the other figures included in this chapter, the whiskers of the boxplots represent the standard deviation and the black dots represent outliers.

In order to confirm whether the trends in the data were statistically significant, the intelligibility accuracy was analyzed with logistic regression with the binomial link function using the *glmer* function from the "lme4" package (Bates et al., 2015) (version 1.1.21) in R (version 3.5.0)(R Development Core Team, 2008). The intelligibility of the words, i.e., whether listeners identified the utterance as the target word or not, was predicted using the theoretically relevant factors: Test (pre-test, post-test), Vowel (/æ, ɛ/), Condition (Control, OnlyExposure, OnlyFeedback, Combined) and their interactions. Random intercepts for Speaker, Listener and Word were also included. No random slopes were included due to convergence issues. The optimizer "bobyqa" was used to improve the model performance.

Table 2 shows the analysis of deviance table for the model fitted to the intelligibility, obtained with the "Anova" function of the "car" package (Weisberg & Fox, 2011) (version 3.0.2) in R (version 3.5.0) (R Development Core Team, 2008). With this function, the Type

II analysis-of-variance table is calculated for the regression model. The output of the model showed a statistically significant interaction between Test, Vowel and Condition, indicating that the accuracy of the native listeners at identifying the words produced by the Dutch learners varied from pre-test to post-test depending on the Vowel and the Condition to which the speakers had been assigned. In order to understand this three-way interaction, the data was split by Vowel.

Table 2. Analysis of deviance table of the model fitted to the Accuracy (whether the vowel was correctly identified as such) as predicted by Test, Vowel, Condition and their interactions.

Predictors	*Chisq*	*Df*	*Pr(>Chisq)*
Test	**37.73**	**1**	**<0.001**
Vowel	**15.69**	**1**	**<0.001**
Condition	4.818	3	0.187
Test:Vowel	**27.16**	**1**	**<0.001**
Test:Condition	**23.51**	**3**	**<0.001**
Vowel:Condition	**324.20**	**3**	**<0.001**
Test:Vowel:Condition	**16.72**	**3**	**<0.001**

Table 3. Output of the model fitted to the Accuracy of the /æ/ tokens as predicted by Test and Condition and their interactions. The intercept represents the pre-test and the Control condition. Significant values are highlighted in bold.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-0.25	0.32	-0.773	0.439
Test(Post)	-0.05	0.06	-0.780	0.435
Condition(OnlyExposure)	0.48	0.32	1.496	0.135
Condition(OnlyFeedback)	-0.29	0.32	-0.906	0.365
Condition(Combined)	0.05	0.33	0.142	0.887
Test(Post): Condition(OnlyExposure)	**0.27**	**0.09**	**3.106**	**0.002**
Test(Post): Condition(OnlyFeedback)	**0.36**	**0.09**	**4.115**	**<0.001**
Test(Post): Condition(Combined)	**0.64**	**0.09**	**7.096**	**<0.001**

The model fitted to the accuracy data for the tokens including the /æ/ vowel, with the pre-test tokens produced by speakers in the Control condition on the intercept, showed no significant simple effect of Test (see Table 3). Thus, no difference was found in the accuracy with which the native listeners identified the tokens containing this vowel between the pre-test and the post-test in the Control condition. In contrast, the significant interactions between Test and the Conditions indicate that there was an effect of Test for the other three conditions. The effect sizes (see Table 3) indicate that the /æ/ tokens were more likely to be identified as such when they were produced in the post-test than in the pre-test in the Only Exposure, Only Feedback and the Combined conditions.

In order to determine whether the pretest-posttest difference in accuracy varied among these conditions, the model was releveled to include the OnlyExposure condition on the intercept (see Appendix 26). The output of the releveled model with the OnlyExposure condition on the intercept showed no significant interaction between Test and the OnlyFeedback condition (β= 0.09, p= 0.305) (see Appendix 26). Thus, the increase in intelligibility in these two conditions was not found to be significantly different. The model revealed a significant interaction between Test and the Combined condition (β= 0.36, p= <0.001). Thus, the pre-test post-test improvement in accuracy in the Combined condition was significantly larger than in the OnlyExposure condition and the OnlyFeedback condition.

Table 4. Output of the model fitted to the Accuracy of the /ɛ/ tokens as predicted by Test and Condition and their interactions. The intercept represents the pre-test and the Control condition. Significant values are highlighted in bold.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	1.24	0.24	5.047	<0.001
Test(Post)	0.00	0.07	0.008	0.994
Condition(OnlyExposure)	-0.40	0.23	-1.681	0.093
Condition(OnlyFeedback)	-0.24	0.24	-1.020	0.308
Condition(Combined)	**-0.47**	**0.24**	**-1.974**	**0.048**
Test(Post): Condition(OnlyExposure)	0.03	0.09	0.317	0.751
Test(Post): Condition(OnlyFeedback)	0.02	0.09	0.236	0.814
Test(Post): Condition(Combined)	0.02	0.10	0.216	0.829

The model fitted to the /ɛ/ tokens, with the pre-test tokens by the participants in the Control condition on the intercept, revealed no significant simple effect of Test (see Table 4). Thus, no evidence was found for the native listeners to be more likely to identify the /ɛ/ tokens correctly when they had been produced in the post-test than in the pre-test in the Control condition. The absence of significant interactions between Test and Condition indicates that this lack of change applied to the other conditions as well. Therefore, no evidence was found of an effect of exposure and feedback triggering the intelligibility of the words containing the vowel /ɛ/.

In summary, Dutch speakers who received exposure and feedback during an interaction produced more intelligible tokens containing /æ/ in the post-test than in the pre-test. The effects of these two factors presented in isolation on the intelligibility of the /æ/ tokens seem to be similar. The presence of the two factors in combination also boosted the intelligibility of the words containing this vowel. In fact, the effect of the combination of the factors on the intelligibility of /æ/ tokens was significantly larger than the sum of the effects of the factors presented in isolation. As for the /ɛ/ tokens, exposure and feedback did not seem to have an effect on the intelligibility of the learners' pronunciation of these tokens, since the accuracy with which native listeners identified words including this sound did not seem to differ from pre-test to the post-test productions in any of the conditions.

4.4. Experiment 2: Nativelikeness Judgements task

The goal of this experiment was to determine whether native listeners perceived an improvement in the nativelikeness of L2 speakers' pronunciation of the critical contrast, /æ, ɛ/, after the L2 speakers had received exposure and feedback during the interaction.

4.4.1. Methods

4.4.1.1. Participants

The same listeners participated in this experiment as those in Experiment 1, except for four participants who could not perform Experiment 2 due to technical issues. All of the participants were rewarded for their time with study credits.

4.4.1.2. Materials

The lists of word tokens used in this experiment were identical to the lists for Experiment 1 (see section 4.3.1.2). Participants were assigned to same list number as in Experiment 1 (list 1: 34 listeners; list 2: 35; list 3: 34).

4.4.2. Procedure

The Nativelikeness Judgement task was programmed using *ROLEG.* Participants performed this task on a computer equipped with a pair of headphones and a keyboard. At the beginning of each trial, the audio recording of a word token was played automatically to the participants' headphones. On the screen, participants could read a sentence stating the word the speaker intended to say, e.g., "The word is **bad**", and a question asking them to rate the nativelikeness of the word. Participants used a 5-point scale, ranging from strongly accented to native-like. In addition, participants were reminded in each trial that the lower extreme of the scale (1) represented a strongly accented non-native pronunciation of the word and that the higher extreme of the scale (5) represented a nativelike pronunciation of the word. Participants pressed the numeric key (from 1 to 5) corresponding to the level of nativelikeness they wanted to assign to the pronunciation of the word token they listened to. Once they judged the word token, the next trial started automatically. After every block of 8 tokens produced by the same speaker, participants saw a screen asking them to take a small break. Participants took approximately 15 minutes to complete this experiment.

4.4.3. Results and discussion

The degree of reliability among listeners who judged the nativelikeness of the tokens produced by the Dutch learners of English was assessed using the intra-class correlation coefficient, which was calculated using the "ICC" function from the "pysch" package (Revelle, 2019) (version 1.9.12.31) in R (version 3.5.0) (R Development Core Team, 2008). Assessing the degree of agreement across lists was possible because of the subset of 13 items that were present in all the three lists of the experiment. The intra-class correlation coefficient, revealed an excellent correlation using the two-way mixed effects model and "multiple rater" unit (ICC3k), kappa = 0.96, p = < 0.001. This excellent agreement was also found among listeners judging the same word lists (List 1: ICC3k kappa = 0.89, p = < 0.001; List 2: ICC3k kappa = 0.92, p = < 0.001; List 3: ICC3k kappa = 0.89, p = < 0.001).

4.4.3.2. Detecting phonetic adaptations in terms of nativelikeness as the result of receiving exposure and feedback

Figure 4 shows the average nativelikeness score listeners provided for the tokens produced by the Dutch speakers separated by Test (pre-interaction and post-interaction), Vowel (/æ, ɛ/) and Condition (Control, Only Exposure, Only Feedback and Combined). The trends in the figure indicate that there were differences between the scores assigned to the two vowels. First, L2 speakers' pronunciation of /ɛ/ is rated overall more nativelike than the pronunciation of /æ/. Second, across the four conditions, the nativelikeness scores for /æ/ seem to vary from pre-test to post-test, while the scores assigned to /ɛ/ tokens appear to remain unchanged from pre-test to post-test.

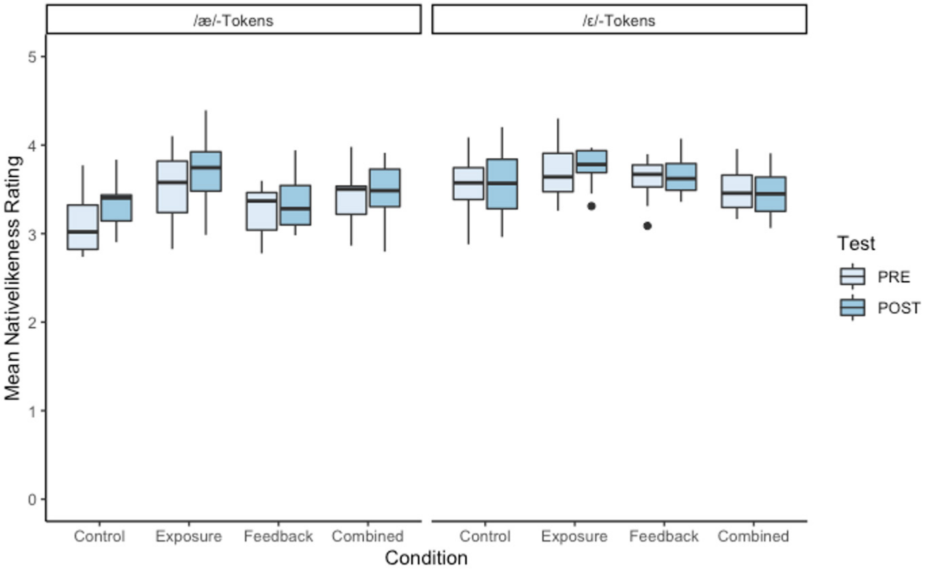


Figure 4. Boxplot representing the distribution of the average nativelikeness score assigned to the learners, grouped by condition, separated by test (in the pre-test in light blue and in the post-test in darker blue) for the vowel /æ/ (on the left) and the vowel /ɛ/ (on the right).

4.4.3.2. Detecting phonetic adaptations in terms of nativelikeness as the result of receiving exposure and feedback

Figure 4 shows the average nativelikeness score listeners provided for the tokens produced by the Dutch speakers separated by Test (pre-interaction and post-interaction), Vowel (/æ, ɛ/) and Condition (Control, Only Exposure, Only Feedback and Combined). The trends in the figure indicate that there were differences between the scores assigned to the two vowels. First, L2 speakers' pronunciation of /ɛ/ is rated overall more nativelike than the pronunciation of /æ/. Second, across the four conditions, the nativelikeness scores for /æ/ seem to vary from pre-test to post-test, while the scores assigned to /ɛ/ tokens appear to remain unchanged from pre-test to post-test.

In order to determine whether the trends reflected in Figure 4 are statistically significant, the nativelikeness ratings provided by the listeners were analyzed using ordinal regression with the *clmm* function from the “ordinal” package (version 2019.12.10) in R (version 3.5.0) (R Development Core Team, 2008).[2] The model structure included the ratings of nativelikeness as the dependent variable and the theoretically relevant factors as fixed effects: Test (pre-test, post-test), Vowel (/æ, ɛ/), Condition (Control, Only Exposure, Only Feedback, Combined), and their interactions. Random intercepts were included for Speaker, Listener and Word. No random slopes were included due to problems with convergence.

Table 5. Output from the model fit to the Nativelikeness scores as predicted by Test, Vowel, Condition and their interactions. The intercept represents the pre-test, the /æ/ vowel and the Control condition. Significant values are highlighted in bold.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Test(Post)	**0.25**	**0.05**	**4.670**	**<0.001**
Vowel(/ɛ/)	**0.80**	**0.20**	**4.137**	**<0.001**
Condition(OnlyExposure)	**0.77**	**0.18**	**4.395**	**<0.001**
Condition(OnlyFeedback)	0.23	0.18	1.333	0.183
Condition(ExposureAndFeedback)	**0.50**	**0.18**	**2.801**	**0.005**
Test(Post): Vowel(/ɛ/)	**-0.19**	**0.08**	**-2.418**	**0.016**
Test(Post): Condition(OnlyExposure)	-0.13	0.07	-1.727	0.084
Test(Post): Condition(OnlyFeedback)	-0.05	0.07	-0.732	0.464
Test(Post): Condition(Combined)	**-0.26**	**0.07**	**-3.501**	**<0.001**
Vowel(/ɛ/): Condition(OnlyExposure)	**-0.50**	**0.07**	**-6.728**	**<0.001**
Vowel(/ɛ/): Condition(OnlyFeedback)	**-0.16**	**0.07**	**-2.225**	**0.026**
Vowel(/ɛ/): Condition(Combined)	**-0.66**	**0.08**	**-8.717**	**<0.001**
Test(Post): Vowel(/ɛ/): Condition(OnlyExposure)	0.11	0.10	1.065	0.287
Test(Post): Vowel(/ɛ/): Condition(OnlyFeedback)	0.10	0.10	0.944	0.345
Test(Post): Vowel(/ɛ/): Condition(Combined)	**0.23**	**0.11**	**2.190**	**0.028**

Table 5 shows the output of the ordinal model fitted to the nativelikeness ratings. This table revealed significant interactions between Test, Condition and Vowel, meaning that the nativelikeness ratings were affected differently depending on the combination of

---

[2] There is little consensus in the literature on how to analyze scale data statistically. In this chapter, we chose the ordinal regression over other methods.

those three factors. In order to interpret this three-way interaction, the data was split by Vowel.

Table 6. Output from the model fit to the Nativelikeness scores assigned to the /æ/ tokens, as predicted by Test and Condition and their interactions. The intercept represents the pre-test productions and the Control condition. Significant values are highlighted in bold.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Test(Post)	**0.25**	**0.05**	**4.715**	**<0.001**
Condition(OnlyExposure)	**0.79**	**0.23**	**3.445**	**<0.001**
Condition(OnlyFeedback)	0.24	0.23	1.061	0.289
Condition(Combined)	**0.51**	**0.23**	**2.208**	**0.027**
Test(Post): Condition(OnlyExposure)	-0.13	0.07	-1.767	0.077
Test(Post): Condition(OnlyFeedback)	-0.06	0.07	-0.828	0.408
Test(Post): Condition(Combined)	**-0.26**	**0.07**	**-3.469**	**<0.001**

The output of the model fitted to the /æ/ tokens with the pre-test productions and the Control condition on the intercept reveals a significant simple effect of Test (see Table 6). This effect indicates that the nativelikeness scores assigned to the /æ/ tokens were significantly higher when they were produced in the post-test than in the pre-test production task in the Control condition (which is the condition on the intercept). The interactions between Test and Condition were not significant for the Only Exposure or the Only Feedback conditions. Therefore, in these two conditions, the post-test tokens of /æ/ were rated higher in nativelikeness than those produced in the pre-test to the same extent as in the Control condition. In contrast, the interaction between Test and Condition was significant for the Combined condition, where exposure and feedback were presented together, which indicated that the effect of Test was different in this condition. In order to interpret this interaction, the model was releveled to include the Combined condition on the intercept (see Appendix 27). The output of the releveled model showed no significant effect of Test ($\beta=-0.01$, $p= 0.906$), meaning that no difference was found between the nativelikeness ratings assigned to the pre-test and post-test productions of

/æ/ in the Combined condition. Note that the condition OnlyExposure and the condition Combined also had simple effects, indicating that the speakers in these conditions happened to have more native-like English accents than the speakers in the Control condition and in the OnlyFeedback condition, which surfaced in both the pre- and the post-test.

Table 7. Output from the model fit to the Nativelikeness scores assigned to the /ɛ/ tokens, as predicted by Test and Condition and their interactions. The intercept represents the pre-test productions and the Control condition.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Test(Post)	0.06436	0.05541	1.161	0.245
Condition(OnlyExposure)	0.26558	0.17820	1.490	0.136
Condition(OnlyFeedback)	0.06475	0.17798	0.364	0.716
Condition(Combined)	-0.18413	0.18161	-1.014	0.311
Test(Post): Condition(OnlyExposure)	-0.01146	0.07516	-0.153	0.879
Test(Post): Condition(OnlyFeedback)	0.04677	0.07417	0.631	0.528
Test(Post): Condition(Combined)	-0.01717	0.07665	-0.224	0.823

The model fitted to the /ɛ/ tokens, with the pre-test tokens and the Control condition on the intercept (see Table 7), showed that the effect of Test was not significant. Thus, the nativelikeness scores assigned to this vowel in the Control condition did not seem to change from the pre-test to the post-test production. The interactions between Test and Condition were not significant either. Therefore, no evidence was found for the nativelikeness scores assigned to /ɛ/ vowels by the native listeners of English to be significantly different from the pre-test to the post-test in any of the four conditions.

In summary, native listeners rated post-test /æ/ tokens overall as more native-like than pre-test tokens of the same vowel, except in the combined Exposure and Feedback condition. The ratings native listeners assigned to /ɛ/ tokens did not seem to change significantly from the pre-test to the post-test in any of the conditions.

4.5. Experiment 3: Automatically Generated Intelligibility scores

Experiment 3 addresses the third research question of the study, which focuses on whether exposure and/or feedback during an interaction can lead to an increase in intelligibility of L2 speakers' pronunciation as indicated by ASA-based technology. Montreal Forced Aligner (MFA) (McAuliffe et al., 2017), a free access forced aligner, with English pre-trained phone models and an adapted dictionary (see section 5.1.2) were used in order to assess the intelligibility of the learners' pronunciation of the critical contrast.

4.5.1. Methods

4.5.1.1. Materials

Experiment 3 uses the same audio fragments as Experiments 1 and 2, but there was no need to order them in lists.

4.5.1.2. Montreal Forced Aligner settings

Montreal Forced Aligner is software used to align audio input with orthographic transcription (McAuliffe et al., 2017). In order to do so, MFA uses acoustic phone models and a pronunciation dictionary with a word-to-phone mapping.

Phone models: Pre-trained English models, based on the LibriSpeech dataset (Panayotov et al., 2015), which are freely available with MFA, were used for this experiment. These models were used because, for this experiment, we hypothesized that they could simulate the sound categories native listeners of English use to identify sounds.

Dictionary: The dictionary used was based on the Carnegie Mellon University (CMU) dictionary (Weide, 1994). Dictionaries used for force-aligning can include different potential transcriptions of the same entry to reflect possible pronunciations of the same word. In the segmental annotation, the forced aligner chooses the transcription which matches the acoustic input best. For the purpose of this study, the CMU dictionary was extended to include possible Dutch-accented pronunciations. For every word present in the materials of the experiment, a possible transcription was added to the dictionary. that reflected a possible Dutch-accented pronunciation of the word. Given that Dutch speakers tend to merge the /æ/ and /ɛ/ vowels into an in-between sound, the "accented" pronunciation consisted of replacing one of the critical vowels with the other. Thus, every keyword included in the materials, e.g., "bad", had both a native-like pronunciation, /bæd/, and a Dutch-accented pronunciation, /bɛd/. The choice between these two transcriptions made by MFA was very similar to the intelligibility task performed by native listeners of English in Experiment 1.

4.5.2. Procedure

The audio recordings of the tokens were force aligned using MFA with the settings described in section 4.5.1.2. Given that the orthographic transcription of each target word had two possible entries in the pronunciation dictionary, the software was forced to choose the pronunciation that matched better with the acoustic signal. Once all the tokens were forced aligned to their transcriptions, the output of MFA was coded similarly to the intelligibility task responses described in Experiment 1 (see section 4.3.2). A binary variable was coded reflecting whether MFA had aligned a token, for instance, "bad", with either the native-like transcription, /bæd/, or the Dutch-accented option /bɛd/.

4.5.3. Results

4.5.3.1. Agreement between automatically-generated and human-provided intelligibility scores

Before analyzing the effect of exposure and feedback on the ASA-based intelligibility scores, we investigated the agreement between intelligibility by MFA and by the human listeners who participated in Experiment 1 (see section 3). For each L2 learner, two overall intelligibility scores were calculated: the first one based on the human listeners' performance in the intelligibility task, and the second was based on MFA's performance in the same task. In order to calculate the human overall intelligibility score for each learner, we calculated the proportion of correctly identified tokens by each of the listeners. A single score was calculated by averaging the mean of scores given by all the human listeners for each learner. The MFA intelligibility score was calculated similarly: for each learner, the proportion of correctly identified tokens by MFA in the intelligibility task was calculated.

Figure 5 presents a scatterplot where every dot represents one learner included in the study. The y axis represents the learner's intelligibility score based on MFA, and the x axis corresponds to the learner's intelligibility score based on the human listeners. The listeners are ordered based on their human intelligibility score on the x axis. This scatterplot seems to show a correlation between the two intelligibility scores: learners who were less intelligible according to the human listeners also seem to be less well understood by the MFA.

In order to determine whether this trend was statistically significant, a Pearson correlation was calculated between the two scores: the human intelligibility score and the MFA intelligibility score. The test revealed a strong positive correlation (r(184)= .73, p = <0.001) between the two variables. This correlation confirms that the overall intelligibility scores based on MFA was strongly correlated with the overall intelligibility based on the human-perceived intelligibility.

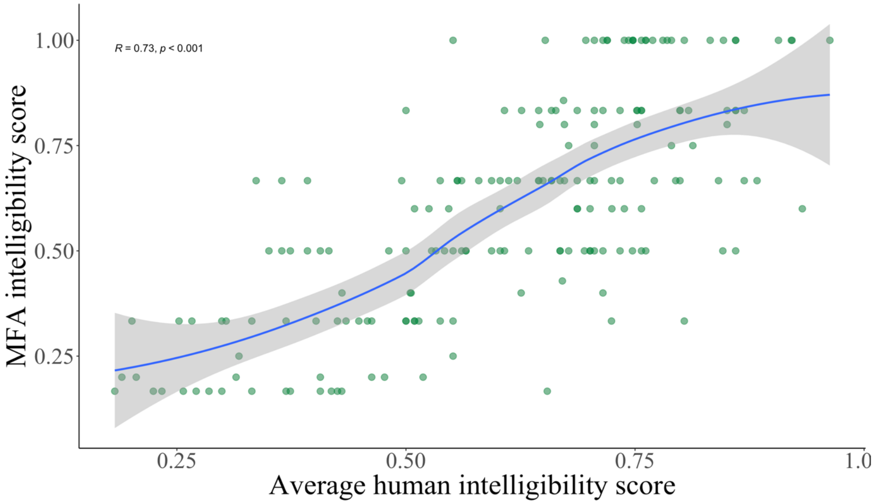


Figure 5. Scatterplot where every green dot represents one learner included in the study and their intelligibility scores based on the MFA's performance (y axis) and based on the human listeners' performance (x axis) in the intelligibility task.

In summary, the performance of MFA in the intelligibility task was found to be comparable to that of the human listeners. When using the responses to the intelligibility task to calculate an overall intelligibility score for each learner, the scores based on the MFA performance were found to be strongly correlated to those based on the human listeners.

4.5.3.2. Detecting phonetic adaptations in L2 learners as the result of receiving exposure and feedback

The goal of this experiment is to determine whether MFA could detect an improvement in the intelligibility of /æ, ɛ/ word tokens produced by L2 learners after they had received exposure and/or feedback during an interaction. Therefore, we analyzed whether the tokens produced by Dutch speakers in the post-test production task were more likely to be annotated by MFA with the native-like pronunciation than with the Dutch-accented pronunciation when the speakers had received exposure and/or feedback.

Figure 6 shows the percentage of word tokens identified by MFA as the containing the vowel intended by the Dutch speaker, summarized per condition, separated by vowel (/æ/ and /ɛ/) and test (pre-interaction and post-interaction). Overall, the intelligibility of MFA seems to be higher for words containing /ɛ/ as compared to words containing /æ/. In addition, the figure suggests a difference depending on the test: the intelligibility in the post-test seems to be overall higher than in the pre-test.

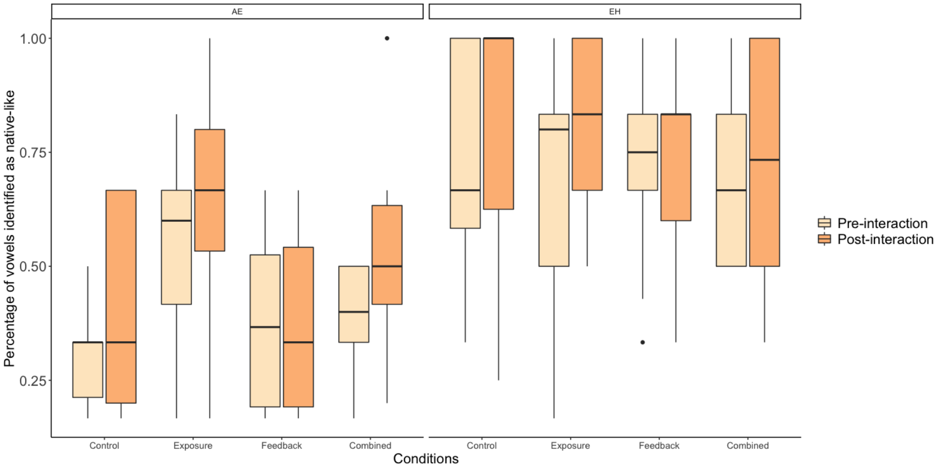


Figure 6. Boxplot representing the distribution of the percentage of vowels identified by MFA in accordance with the speaker's intention, separated by condition (Control, OnlyExposure, OnlyFeedback and Combined), vowel (/æ/ and /ɛ/) and test (pre-interaction and post-interaction).

In order to determine whether the trends in Figure 6 were statistically significant, the MFA-generated intelligibility scores were analyzed using logistic regression with the binomial link function using the glmer function from the "lme4" package (version 1.1.21) in R (version 3.5.0). The intelligibility of the words, i.e., whether they were identified by MFA as the target word or not, was predicted using the theoretically relevant factors: Test (pre-test and post-test), Vowel (/æ/ and /ɛ/), Condition (Control, Only Exposure, Only Feedback, Combined) and their interactions. The structure of the fixed effects was im proved by removing interactions and simple effects that were not considered statistically significant, with p-values higher than 0.05, if their removal improved the model AIC value. Random intercepts for Speaker, Listener and Word were also included. Due to convergence issues, no random slopes could be added. The optimizer "bobyqa" was used to improve the model performance.

Table 8. Output of the model fitted to the MFA intelligibility scores (whether the vowel was correctly identified as such by MFA) as predicted by Test, Vowel, Condition and their interactions. Significant values are highlighted in bold.

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-1.26	0.47	-2.688	0.007
Test(Post)	**0.31**	**0.14**	**2.155**	**0.031**
Vowel(/ɛ/)	**2.73**	**0.66**	**4.163**	**<0.001**
Condition(OnlyExposure)	**1.07**	**0.30**	**3.555**	**<0.001**
Condition(OnlyFeedback)	0.25	0.30	0.837	0.403
Condition(Combined)	0.59	0.31	1.934	0.053
Vowel(/ɛ/):Condition(OnlyExposure)	**-1.37**	**0.42**	**-3.244**	**0.001**
Vowel(/ɛ/):Condition(OnlyFeedback)	-0.60	0.42	-1.426	0.154
Vowel(/ɛ/):Condition(Combined)	**-1.02**	**0.43**	**-2.378**	**0.017**

Table 8 shows the output of the final model fitted to the MFA intelligibility data. It shows a significant simple effect of Vowel, indicating that tokens including the /ɛ/ vowel were significantly more often correctly identified than those including the /æ/ vowel. There is also a significant simple effect of Test. This effect indicates that the intelligibility of the post-test tokens was significantly higher than the intelligibility of the pre-test tokens. The lack of significant interactions between Test and Condition, Test and Vowel, or Test, Vowel and Condition, indicates no evidence was found for this increase in the intelligibility to be different depending on the vowel included in the token or the condition to which the learners had been assigned.

4.6. General discussion

The current study investigated whether receiving exposure and feedback during an interaction affects L2 learning via phonetic accommodation. Experiment 1 shows that these two factors fostered an improvement in the intelligibility of Dutch speakers' pronunciation of the word tokens including the /æ/ vowel, as perceived by native listeners of the L2. Experiment 2 shows that L2 learners adopted a more native-like pronunciation of the /æ/ vowel just by participating in the conversation, except when during the conversation they received both exposure to the critical vowel and feedback on their own pronunciation. In that latter case, the combination of exposure and feedback hindered the overall improvement in nativelikeness. Similarly, Experiment 3 shows that ASA-based technology detected an improvement in the intelligibility of the learners' pronunciation just as a result of taking part in the conversation. In this experiment, however, the presence of exposure and feedback, in isolation or combination, did not have an effect on the intelligibility at all.

4.6.1. Exposure

Previous research investigating the effect of exposure in L2 learning and phonetic accommodation has found that only being exposed to the interlocutor's pronunciation can lead speakers to adapt their own pronunciation to match the interlocutor's speech (e.g., Babel, 2012; Babel et al., 2013; Berry & Ernestus, 2018; Delvaux & Soquet, 2007; Gessinger et al., 2019; Goldinger, 1998; Hwang et al., 2015; Namy et al., 2002; Pardo et al., 2012). Our study adds to this literature that in our experimental design, we combine the careful phonetic control, implemented by using pre-recorded speech, together with a highly interactive and spontaneous setting. This set-up makes the effect of exposure as a trigger for phonetic accommodation found in this study easier to generalize to real-life everyday interactions L2 learners may encounter.

The results of this study show that the presence of controlled exposure during a spontaneous interaction with a native speaker affected the intelligibility and nativelikeness of the L2 speakers' pronunciation differently. Receiving exposure during the interaction led to more intelligible pronunciations of the word tokens including the /æ/ vowel in the post-test as perceived by the native listeners (Experiment 1). However, this improvement in intelligibility was not detected by the ASA system (Experiment 2). Exposure to the native-like pronunciation did not seem to booster or hinder the nativelikeness of the L2 learners.

The lack of an effect of exposure on the nativelikeness and the MFA intelligibility could be related to the room for improvement. The

L2 learners in this study displayed an overall improvement in these two perceptual dimensions, nativelikeness and MFA intelligibility, just by taking part in the interaction. Therefore, this overall improvement may leave no room for improvement as the result of receiving exposure.

4.6.2. Feedback

Feedback has been reported to be a trigger for L2 sound learning: providing learners of an L2 with feedback on their pronunciation can lead them to adopt a more native-like pronunciation (Aliaga-García & Mora, 2008; Burnham et al., 2010; Dlaska & Krekeler, 2013; Kartushina et al., 2015; Saito & Lyster, 2012). Research from the perspective of phonetic accommodation has also found that speakers adapt their pronunciation after receiving feedback from a human interlocutor (Burnham et al., 2010). The current study combines these two perspectives by studying the effect of feedback on phonetic accommodation in situations involving L2 speakers. Crucially, by means of using pre-recorded speech avoiding the presence of exposure to the critical contrast, the effect of feedback on these sounds is examined separately from the effect of receiving exposure.

The results of this study in terms of intelligibility follow the literature on feedback in L2 learning, in classroom settings and with computer programs, where feedback is found to be an effective tool for L2 teaching (e.g., Adams et al., 2011; Dlaska & Krekeler, 2013; Lee & Lyster, 2017; Nassaji, 2009; Saito & Lyster, 2012). Feedback implying misunderstanding from the interlocutor was found to make speakers produce more intelligible /æ/ tokens to the native ear (Experiment 1). The effect of feedback was found to be similar to that of exposure in isolation. Interestingly, this increase in intelligibility was not detected the ASA-technology (Experiment 3). The nativelikeness of the pronunciation was not found to be affected by the presence of feedback during the interaction (Experiment 2). Similarly to Exposure, the lack of an effect of feedback on the nativelikeness and the MFA intelligibility could be related to the room for improvement.

This study expands the scope of the study of the effectiveness of feedback to a currently understudied context for language learning: interactions in the L2. Unlike previous literature investigating the effect of feedback on pronunciation adaptations both from an L2 acquisition perspective and from the field of phonetic accommodation, the current study discerns the effect of feedback from the possible benefits speakers can obtain from receiving exposure to the target pronunciation of the critical sounds. By using pre-recorded speech, we ensured the learners in our study did not receive exposure at all to the interlocutor's pronunciation of the critical contrast. Results confirm that, even when the learners did not receive exposure to the target pronunciation of the critical sounds, receiving feedback led them to produce these vowels in a more intelligible way, thus, feedback in isolation proved to be an effective tool in L2 learning in conversation.

4.6.3. The combination of Feedback and Exposure

The results of this study show that the combination of exposure and feedback affected intelligibility and nativelikeness differently. In Experiment 1, the combination of the two factors led to an improvement on the intelligibility perceived by the native listeners. This improvement was found to be significantly larger than the effect of the two factors in isolation on the human intelligibility. Interestingly, the ASA-technology failed to find this difference: in Experiment 3, the combination of the two factors triggered an increase in intelligibility that was not found to be different from the increase boosted by the presence of exposure in isolation.

In terms of nativelikeness, the combination of the two factors seemed to hinder nativelikeness (Experiment 2). The learners who received both exposure and feedback during the interaction did not show the overall increase in nativelikeness that the learners in the other three conditions displayed, including the Control condition learners. Therefore, although the two factors in isolation did not seem to affect the overall increase in nativelikeness, the combination of exposure and feedback seemed to hinder phonetic accommodation to the interlocutor's pronunciation.

These results have consequences for the study of phonetic accommodation in general. The differences found between the effects of exposure and feedback in isolation and in combination on both the human intelligibility and the nativelikeness call for a more detail-oriented study of the mechanisms underlying phonetic accommodation. In particular, this study sheds some light on the importance of controlling the presence of exposure when investigating how phonetic accommodation may be affected by any other factor, for instance, feedback. Failing to disentangle these two factors in the experimental design of this study would have led to incomplete conclusions about the effect of feedback on the perceived phonetic accommodation from L2 learners to a native speaker.

4.6.4. Differences between /æ/ and /ɛ/

The results of his study show an asymmetry between how the two vowels investigated –the English /æ, ɛ/ contrast – were affected by L2 speakers aligning to the interlocutor's pronunciation. While /æ/ underwent drastic changes in intelligibility (Experiments 1 and 3) and nativelikeness (Experiment 2); /ɛ/ was not affected in terms of nativelikeness or human-perceived intelligibility at all. Only ASA-technology (Experiment 3) seemed to detect changes in this vowel, which overall became more intelligible in the post-test. There could be two explanations for this asymmetry between the two vowels.

The first explanation is based on the room for improvement each vowel offered in the baseline pronunciation of the L2 speakers. Previous research on the Dutch-accented pronunciation of English has reported that Dutch speakers tend to merge these two vowels into an in-between sound, which is closer to /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006). In fact, the acoustic analysis reported in Chapter 3 also revealed that, in Dutch learners' vowel space, English /ɛ/ was acoustically closer to the native-like pronunciation than /æ/. The previous literature is mainly based on acoustic analyses of the vowels. The current study shows that the asymmetry between the vowels also holds from a perceptual perspective. The accuracy with which the native listeners and the ASA-technology identified the vowel intended by the L2 learners was overall higher for /ɛ/ than for /æ/. This tendency confirms that /ɛ/ is closer to the native-like pronunciation of the vowel and that Dutch-accented /æ/ is often interpreted as English /ɛ/. Similarly, the nativelikeness ratings were also higher for /ɛ/ than for /æ/. Given that the baseline pronunciation of /æ/ was less intelligible and less native-like than the pronunciation of /ɛ/, and that the intelligibility and nativelikeness of /ɛ/ was close to ceiling, the /æ/ tokens could have been affected more in the process of accommodation simply because this vowel offered more room for improvement than the baseline pronunciation of Dutch-accented English /ɛ/ did. This explanation would follow previous literature suggesting that phonetic features that are perceptually more different between two interlocutors are more likely to be subject to accommodation (MacLeod, 2014).

The second possible explanation could be related to the dynamics of the puzzle-solving task used in the production experiment. As explained in section 4.2, during the Code Breaker trials, the learners had to instruct the confederate to click on a word as part of the game. In a total of twelve critical trials, learners had to produce a word containing the /æ/ sound. The /ɛ/ vowel, in contrast, was never included as the word the learners had to instruct the confederate to click on. Therefore, during the game, /æ/ received substantially more attention than /ɛ/, making it perhaps more susceptible to change based on exposure and feedback. For instance, the learners could have interpreted the feedback they received to be directly aimed at their pronunciation of /æ/, and they may not think it was necessary for them to change the pronunciation of the /ɛ/.

4.6.5. Exposure and feedback trigger phonetic accommodation in different perceptual dimensions

This study shows that the choice of which perceptual dimension is used to measure phonetic accommodation is crucial because not all dimensions are affected equally. In this study the intelligibility and the nativelikeness of the pronunciation of L2 speakers was affected differently as the result of receiving exposure and/or feedback. Human intelligibility was fostered by exposure and feedback, in isolation and combination; nativelikeness seemed to increase just by participating in the interaction but was hindered by the combination of the two factors.

This lack of consistency between the results of the two perceptual tasks performed by the human listeners suggests that listeners may relay in different acoustic cues available in order to judge the intelligibility and the nativelikeness of L2 pronunciation. L2 speakers may be adapting some of those cues but not all of them equally when aligning to the interlocutor. Even if listeners took into account the same cues for the intelligibility and nativelikeness, the adjustments learners made to their speech to make the pronunciation of the two sounds more intelligible may also make the sounds less native-like to native ears. The results of the Combined condition, where speakers were judged to be more intelligible but less native-like, seem to point in this direction. Further research comparing the perception results to measurements of the acoustic properties of the pronunciation of these vowels, for instance, duration and the intonation used, could help disentangle whether the same acoustic cues with the same or different weights were used by the listeners performing each perception task.

The intelligibility task is a direct method to collect information about how intelligible the pronunciation of a sound actually is. The listeners have only two options from which to choose: they had to identify a word between only two possible candidates. The nativelikeness task, on the other hand, involves more explicit judgement from the listeners. They are presented with the information and asked to evaluate the pronunciation using 5 steps on a scale. In this study, it seems like more implicit methods, such as intelligibility judgments, may give a more detailed and less subjective representation of listeners' judgement on the pronunciation of L2 speakers.

4.6.6. Differences in human and ASA-based intelligibility scores

The results of this study show differences in the intelligibility of non-native pronunciation as perceived by the human listeners and the ASA-based technology. MFA's performance in the intelligibility task was found to be similar to that of the human listeners. In particular, the overall intelligibility scores based on MFA, i.e., the proportion of tokens per speaker correctly identified by MFA in the intelligibility task, were also found to be strongly correlated to the overall intelligibility scores based on the human listeners' performance in the same task. These results should be interpreted with caution, given that they are based on the intelligibility of only two sounds of a specific L2 speaker population. However, the similarity between these scores seems to suggest that, in order to evaluate the intelligibility of L2 speakers, ASA systems could provide intelligibility scores that are similar to those collected from a vast number of native listeners but in a more efficient way.

Using the MFA responses to investigate how the non-native speakers' intelligibility was affected by the presence of exposure and feedback seemed to show substantial differences between the human judgements and MFA-generated ones. These differences arose even though the methodology used to collect the intelligibility was identical: in both cases the intelligibility is the product of an agent, a native speaker or MFA, identifying a word between two possible candidates. Human listeners only perceived an improvement in the intelligibility of /æ/ as the result of both exposure and feedback, together and in isolation. MFA, on the other hand, detected an overall improvement in the intelligibility of the two vowels, /æ/ and /ɛ/, regardless of the presence of exposure and feedback. Consequently, although the overall intelligibility scores were similar between MFA and the human listeners, MFA seems to be less sensitive to detect pronunciation adaptations than human listeners are. Thus, the use of ASA systems to replace human listeners' perception of pronunciation adaptations needs to be further investigated and fine-tuned.

One possible explanation for the different patterns found in

Experiment 1 and Experiment 3 could be related to uncontrolled variability among the human listeners. The large pool of participants recruited for Experiment 1 inevitably brought variability to the data (see the kappa values reported in section 4.3.3.1). This variability could be driven by some of the human listeners behaving very differently from other participants in the pool. In order to prevent this variability from affecting the results, the lists of materials used in Experiment 1 were designed to ensure every listener listened to all the participants in all four conditions. Furthermore, in order to account for this variability in the analysis, the regression model used to analyze the data from Experiment 1 (described in section 4.3.3.2) included a random intercept for listener. Unfortunately, due to convergence issues, no random slopes for listener by test, could be added to this model. That is, the variability in the listeners' responses to pre-test and post-test responses in the identification task was not accounted for. In order to check whether the effects found in the output of the model (described in section 4.3.3.2) were driven by a small sample of the participants, we proceeded to repeat the analysis five times with random subsets of the listeners. Each of the five samples of listeners included 80% of randomly selected listeners out of the total pool of participants. The output of the five models revealed the same results as the original model. Thus, the significant results found in the model are not likely to be driven by small samples of listeners behaving in an unusual way.

The room for variation could have also allowed human listeners and MFA to rely on a slightly different sets of acoustic cues in order to identify the words, or to use the same set of acoustic cues with different weights. The cues adjusted seemed to be taken into account by the human listeners when they were performing the identification task. However, MFA could be using a different set of acoustic cues, not including the one adjusted as a result of receiving exposure and/or feedback. Alternatively, MFA, in the process of annotating a segment, could be using the same acoustic cue that was adjusted, but perhaps with a lighter weight. Therefore, the adjustments resulting from exposure and feedback would go unnoticed by MFA. Alternatively, the different patterns of results could also be related to the statistical power available in each analysis. In the human intelligibility analysis, each word token is identified by 35-37 listeners (resulting in 41,518 data points), whereas in the MFA intelligibility analysis every token is only represented by one single observation (resulting in 1,011 data points). Thus, the lack of significant effects in the model with the MFA data could also be due to low statistical power.

Most importantly, this study shows that more research is needed in order to understand whether native listeners' judgements on the intelligibility of L2 learners can be replaced by ASA-based technology. Typically, the use of ASA-based technology to evaluate pronunciation is based on the comparison between human-human and human-ASA degrees of agreement. Unlike other studies, our comparison was not based on the agreement between ASA-technology and a small number of professionally trained annotators. Instead, the performance of MFA judging the utterances was contrasted with 107 native listeners of the language performing an identical task. Further research should investigate the role of the statistical power in the differences between MFA and human listeners, for instance, by increasing the number of tokens that the comparison is based on. Furthermore, future studies should investigate the underlying factors enabling human native listeners and ASA-based systems in the decision-making process of identifying vowels produced by non-native speakers. Identifying which acoustic cues are relevant in this process and with which weights they are used could help understand the differences between human-based and MFA-based intelligibility.

4.6.7. Differences between perceptual and acoustic analyses

Following previous literature (e.g., Pardo et al., 2013), the current study calls for the implementation of both acoustic and perceptual measurements in the study of phonetic accommodation. The current study, using perceptual measurements, shows that the presence of exposure and feedback in an interaction can lead non-native speakers to adapt their speech in ways that are perceivable, at least in terms of intelligibility, by the human native ear. In the context of phonetic accommodation, where these adaptations could be implemented to ease the process of understanding the message for the interlocutor, knowing whether listeners can perceive the adaptations is crucial.

Crucially, neither of the three perceptual dimensions presented in the three experiments of this study (human intelligibility, nativelikeness and MFA intelligibility) showed a one-to-one mapping with the patterns of results from the acoustic analyses presented in Chapter 3. The acoustic analyses investigating the effect of exposure and feedback on the height and frontness of the critical vowels showed that exposure and feedback in isolation, but not in combination, could trigger adaptations on these two dimensions of the sounds. Human intelligibility was fostered by the presence of exposure and feedback. Nativelikeness was hindered by the combination of the two factors together and MFA-based intelligibility was not found to be affected by exposure and feedback at all.

The differences between the acoustic analyses and the conclusions from the perceptual experiments presented in this study could be related to the scope of the different analysis. Whereas the acoustic analyses focused on two different acoustic dimensions, namely, the height and frontness of the vowels; the perceptual measurements took into account every phonetic cue available for the listeners in the moment of judging the pronunciation of the learners. Thus, the perceptual measurements could be reflecting changes in other phonetic cues that were left outside of the scope of the formant analysis, for instance, vowel duration or intonation.

Alternatively, the differences between the acoustic analyses of the formants and the perceptual dimensions could also be linked to the physical limitations of human perception. Some acoustic adaptations reported in Chapter 3 could have been unperceivable to the human ear when providing the intelligibility and the nativelikeness judgements on the pronunciation of the vowels by Dutch learners. In the case of the nativelikeness, it could be the case that the adjustments the learners made to their pronunciation did not translate into more native-like pronunciations to a native ear.

4.6.8. Phonetic accommodation in L2 speakers

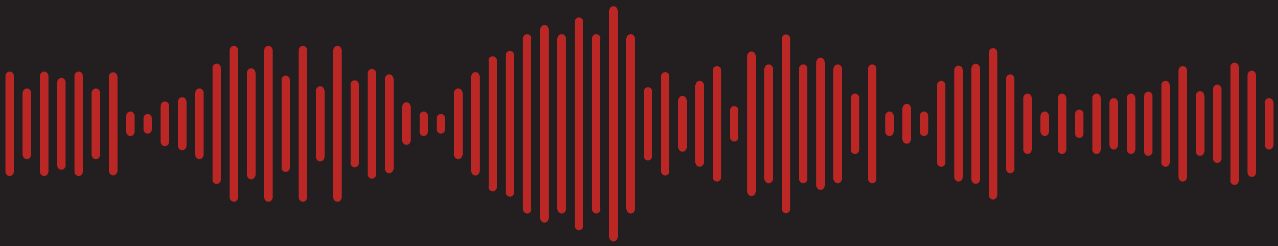
Although accommodation was first assumed to be less likely to take place when speakers use an L2 than when they use their L1 (Costa et al., 2008b), the current study contributes to the recent literature documenting the presence of phonetic accommodation in interactions including non-native speakers (Berry & Ernestus, 2018; Hwang et al., 2015; Lewandowski & Nygaard, 2018; Rojczyk, 2013; Simonet, 2014). In addition, this study also offers some explanation as to which factors seem to be driving the presence of accommodation in L2 speakers.

In particular, here we show that exposure to the native-like pronunciation of the L2 sounds can trigger adaptation by L2 learners. This finding implies that, although L2 learners use more cognitive resources when using an L2 than when using the L1 (Cenoz & García-Lecumberri, 1999), they still have enough resources free to process the exposure they receive from the native interlocutor and to learn from the exposure in the course of the interaction. In addition, in this study, the exposure consisted of listening to an English contrast with which Dutch speakers tend to struggle in terms of perception (Díaz et al., 2012). Thus, exposure seems to be beneficial towards phonetic accommodation even when it includes difficult sound contrasts.

Furthermore, this study shows that phonetic accommodation can also take place as the result of receiving feedback implying misunderstanding during the interaction. Thus, phonetic accommodation in L2 learners was also found to be triggered as a response to ensure understanding in an interaction. Thus, this study calls for the control of phonetic exposure when investigating other factors, such as feedback in the study of phonetic accommodation. Overall, these findings imply that phonetic accommodation in L2 speakers need not be only the result of mere exposure but can also be the result of speakers' effort to ensure understanding. Thus, this study provides evidence for an explanation of phonetic accommodation combining both a speaker-driven perspective, i.e., L2 speakers align as the result of priming, and a listener-driven perspective, i.e., L2 speakers align in order to ensure understanding from the interlocutor.

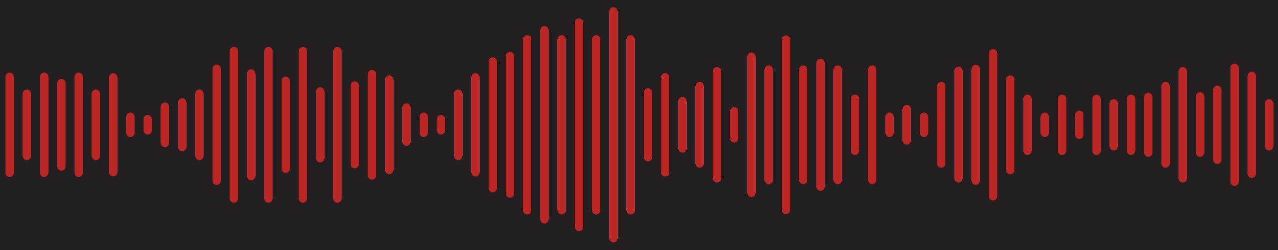
4.6.9. Conclusion

This study investigated the effects of exposure and feedback on the pronunciation of L2 speakers from a perceptual perspective, including the intelligibility and nativelikeness scores of the L2 speakers' pronunciation by native listeners of the L2 and the intelligibility as detected by ASA-based technology. The findings of the study show that phonetic accommodation in L2 speakers can be triggered both by exposure and feedback. The adaptations triggered by these factors may make speakers sound more intelligible but not necessarily more native-like. This study also investigated the use of ASA-systems to evaluate the intelligibility of non-native speakers. The performance of an ASA-system, MFA, in an intelligibility task was found to be similar to that of human listeners, in that the intelligibility scores generated for each learner based on ASA's capacity to identify the vowel produced by non-native speakers were strongly correlated to the intelligibility scores based on the human listeners' vowel recognition. Nevertheless, we failed to find a direct replication of results when the intelligibility task was performed by humans or by MFA. The results from the three experiments presented in this study do not map one-to-one with the acoustic analysis in Chapter 3 This highlights the importance of using perceptual measurements in phonetic accommodation research to determine whether the adaptations can be perceived by the interlocutor. In addition, the results of this study also emphasizes that different perceptual dimensions, such as intelligibility and nativelikeness, can be affected differently when L2 speakers align to a native speaker. The adaptations that may improve the intelligibility of a non-native speaker may, at the same time, hinder their nativelikeness. Finally, this study documents the presence of phonetic accommodation in L2 speakers and shows how aligning behavior at the level of the pronunciation can be both a listener-driven and a speaker-driven phenomenon. Studying how phonetic accommodation is perceived by the interlocutor is crucial in order to understand the motivations behind this phenomenon.



Chapter 5

The effect of proficiency in L2 sound learning through phonetic adaptation in human-human and human-computer interactions



ABSTRACT

This study investigates the role of L2 proficiency on phonetic accommodation, i.e., the process by which a speaker adapts their speech to the interlocutor. L2 proficiency may affect accommodation as it determines a L2 speaker's available cognitive resources (Berry & Ernestus, 2008). The goal of the study is to examine how the level of L2 proficiency modulates phonetic accommodation triggered by exposure to the native pronunciation of the L2 and/or feedback on the L2's speakers' pronunciation. Since phonetic accommodation is dependent on the type of interlocutor (Branigan et al., 2010), the effect of proficiency on accommodation is investigated in human-human and human-computer interactions. We analyzed acoustic data on the English sound contrast /æ, ɛ/ from native speakers of Dutch interacting in English with either a human confederate or a computer. The proficiency measured was based on a variety of self-rated oral proficiency measures, combined in the first component of a principal component analysis. This component was correlated to native speakers' judgements on the nativelikeness of the L2 learners' speech. The results show that in the human-human setting, higher proficiency speakers showed more accommodation than lower proficiency speakers but only when they did not receive exposure to the sounds under investigation or feedback on those sounds during the interaction. In the human-computer setting, in contrast, lower proficiency speakers showed more adaptation than higher proficiency speakers, but only when they received both exposure and feedback on the sound contrast during the interaction. In conclusion, this study provides evidence that phonetic adaptation in non-native speakers is driven by communicative success in interactions with computers.

5.1. Introduction

Imagine you are a non-native learner of English interacting with a native speaker of English. You may notice that your pronunciation of English is accented and, during the conversation, you may attempt to change the way you produce certain sounds to bridge those differences and sound more similar to the native speaker. It could be the case that your level of proficiency in the L2 plays a role in how much you can adapt your pronunciation. If you are a highly proficient learner of English, you may find the conversation to be overall less challenging and you may feel more comfortable and capable to perform adaptations to your pronunciation. If your proficiency level is low, your L2 pronunciation may be more accented and you may feel more motivated to adapt you pronunciation in order to ensure that the native speaker understands what you are saying. In both cases, your readiness to adapt your pronunciation in the course of the conversation may be different if your conversation partner is not a native speaker of English but an interactive computer, in the style of a virtual assistant, using English. In fact, the effect of proficiency on how much you adapt your speech may differ if you are interacting with a computer.

The current study investigates phonetic accommodation, the process by which a speaker may adapt their speech to the interlocutor, in contexts involving non-native speakers. The goal of this study is to determine whether proficiency facilitates or hinders L2 speakers' phonetic adaptation in human-human and human-computer settings.

5.1.1. Phonetic adaptation in conversations

Speakers tend to adapt their speech depending on the specific circumstances of the interaction. For example, speakers adapt their pronunciation in adverse conditions, such as a noisy environment (Lombard, 1911) or when interacting with a child (Fernald et al., 1989). Furthermore, speakers can also adapt their speech when they interact with speakers of a different linguistic variety (Babel, 2010; Pardo et al., 2012). Usually, the result of this latter type of adaptations is that the speaker sounds more similarly to the interlocutor. This phenomenon, known as accommodation (Giles & Ogay, 2007), *alignment* (Pickering & Garrod, 2004), or entrainment (Brennan & Clark, 1996; Levitan & Hirschberg, 2011), has been found to affect different levels of speech, such as the syntactic, lexical, or phonetic level[3].

Most of the literature documenting accommodation in interaction focuses on speakers of different varieties of the same native language (L1) (for example, in English: Babel, 2010, 2012; Pardo et al., 2012, among others; in Spanish see (O'Rourke & Potowski, 2016; Romera & Elordieta, 2013; Troncoso-Ruiz & Elordieta, 2018). These studies show that speakers can adjust features of the variety of a language they speak when they interact with speakers of a different variety of the same language. For example, the intonational contours and the pronunciation of specific segments have been found to be subject to phonetic accommodation (e.g., Babel, 2010, 2010; Cohn et al., 2019; Eijk et al., 2019; Gessinger et al., 2019; Pardo et al., 2012; Romera & Elordieta, 2013, among others).

Although the main body of literature regarding phonetic adaptation has focused on this phenomenon in interactions with native speakers of the same language, the scope has recently broadened to study new contexts, such as non-native speakers (e.g., Berry & Ernestus, 2018; Gnevsheva et al., 2021; Hwang et al., 2015; Kim et al., 2011; Liu & Johnson, 2017). These new contexts help improve our understanding of this phenomenon and, in the case of non-native situations, it may also help unveil how L2 learning works in interactions. In addition, the rise of technology in the field of interactive computer assistants in the 21st century has also led to the broadening of the study of phonetic accommodation to contexts where speakers interact with interactive computers (e.g., Branigan et al., 2010), providing us with insights on how differently people adapt their speech to humans and to computers. The present study aims to combine the new contexts by investigating what drives phonetic adaptation in L2 settings and how comparable L2 phonetic adaptation to humans and to computers is.

---

[3] Notice that the different terms use to discuss speech adaptation –accommodation, alignment and entrainment– come with slightly different connotations. Whereas accommodation is understood as any speech adaptation based on current exposure during the conversation or previous knowledge, alignment is generally understood as repetitive behavior triggered by explicit exposure during an interaction. Here, we focus on phonetic adaptation understood as accommodation.

5.1.2. Phonetic adaptation in non-native contexts

Interactions including a non-native speaker and a native speaker can be interesting contexts to study phonetic adaptation because the gap between the interlocutors' pronunciation is generally greater than in interactions involving speakers of the same language. The non-native interlocutors face greater room for adaptation. The differences between non-native and native speech usually arise from the lack of a one-to-one mapping between the sound systems of the non-native speaker's L1 and the L2. Non-native speakers who learn a new language tend to map the sounds of their L1 onto the L2 sounds (Flege, 1993). When the L1 and the L2 sounds do not completely overlap, mismatches in the pronunciation of given segments could arise (Flege, 1993). For example, the English /æ, ɛ/ vowel contrast, as in "bad"-"bed", has been found to be difficult for speakers of Chinese (Wang & van Heuven, 2006; Zhang & Yin, 2009), German, and Dutch (Bohn & Flege, 1990, 1992; Cutler et al., 2004). The sound systems of these three languages do not include a vowel contrast that is similar to the English "bad"-"bed" pairing and, as a result, these sets of learners struggle both perceiving (Broersma & Cutler, 2011) and producing (Elsendoorn, 1985; Wang & van Heuven, 2006) these two sounds. Non-native speakers who engage in an interaction with a native speaker could attempt to adapt their pronunciation of these difficult contrasts to sound more similar to the native speaker they interact with. The phonetic adaptation taking place in the course of an interaction could then lead to L2 sound learning.

Indeed, phonetic adaptation has been documented in non-native contexts. For example, in situations involving non-native speakers interacting with a native speaker of the L2, non-native speakers have been found to adapt their pronunciation of certain sounds to sound more native-like (e.g., Hwang et al., 2015; Kim et al., 2011, Chapter 3 and 4). In this type of situation, the phonetic adaptation can also go in the opposite direction, i.e., native speakers can also show accommodation to non-native speakers' pronunciation (Lewandowski & Nygaard, 2018). In addition, not only can non-native speakers adapt their pronunciation to native speakers of the L2, but also to other non-native speakers with different language backgrounds. For example, Berry and Ernestus (2018) found that Spanish learners of English accommodated their pronunciation of an English contrast with which they tend to struggle, the English /ɪ, iː/, when interacting with Dutch learners of English. The current study aims to shed some light on which factors drive phonetic adaptation in non-native speakers of a language.

5.1.3. The role of proficiency in non-native phonetic adaptation

The findings from studies investigating phonetic adaptation suggest that many factors may play a role in the degree of adaptation that a speaker can show in an interaction. Two important factors are exposure and feedback. The amount and type of exposure a speaker receives to the interlocutor's speech during the conversation has been argued to be a relevant factor in the triggering of phonetic accommodation (Babel et al., 2013; Babel & Bulatov, 2012; Berry & Ernestus, 2018; Costa et al., 2008; Delvaux & Soquet, 2007; Hwang et al., 2015; Pardo, 2006). Similarly, feedback implying misunderstand by the interlocutor may also be relevant (Burnham et al. 2010; Gessinger et al. 2019). In studying these two factors in non-native contexts, taking into account the L2 proficiency level of the speaker is crucial, since an L2 speaker's proficiency level could modulate how effective exposure and feedback are in triggering phonetic accommodation (Berry & Ernestus, 2018; Kim et al., 2011).

5.1.3.1. Does proficiency foster accommodation triggered by exposure?

In an interaction, the speaker receives exposure to the interlocutor's speech. This exposure has been argued to be the source of phonetic accommodation (Babel et al., 2013; Babel & Bulatov, 2012; Berry & Ernestus, 2018; Delvaux & Soquet, 2007; Hwang et al., 2015; Pardo, 2006). In fact, exposure has been found to trigger phonetic adaptation even when presented as passive ambient sound, as in Delvaux and Soquet (2007). The effect of exposure on phonetic adaptation has been found both in native (Babel et al., 2013; Pardo, 2006) and non-native speakers (Berry & Ernestus, 2018; Delvaux & Soquet, 2007; Hwang et al., 2015).

The effect of exposure on accommodation has been explained in terms of priming. When a speaker receives exposure to a specific cue during an interaction, for instance a word or an allophonic variant of a sound, the higher level of activation of this cue will facilitate the future retrieval and use of the same cue. This perspective defines phonetic accommodation as a speaker-driven phenomenon: speakers accommodate because using a highly activated cue is more efficient for themselves than using a cue with a lower level of activation (Costa et al., 2008; Pickering and Garrod, 2004).

From this perspective, non-native speakers have been theorized to be less capable of adapting their speech during the course of an interaction than native speakers (Costa et al., 2008). If phonetic accommodation is triggered by priming, this phenomenon should be subject to cognitive resources available to the speaker when interacting with an interlocutor (Costa et al., 2008). In order to be primed by exposure to the interlocutor's pronunciation, speakers would need to have enough cognitive resources available to pay close attention to the fine-grained details of the input they receive (Costa et al., 2008). Using an L2 is an activity that usually requires a higher cognitive load than using one's native language. Retrieving L2 lexical items and their native-like pronunciation, among other processes, can be an arduous task (Lecumberri et al., 2010). L2 speakers could thus be assumed to have fewer cognitive resources available to pay close attention to the characteristics of the input than native speakers of the L2. Thus, L2 speakers would be less likely to accommodate because they would be less susceptible to the priming due to the cognitive resources available (Costa et al., 2008).

If the lack of cognitive resources available does indeed hinder phonetic accommodation, the L2 proficiency level of the non-native speaker could be expected to modulate the degree of phonetic accommodation triggered by exposure too (Berry & Ernestus, 2018).

Speakers with a high proficiency level in the L2 tend to process L2 speech more fluidly than low proficiency speakers (Archila-Suerte et al., 2012; Berry & Ernestus, 2018; Bohn & Flege, 1992; Flege et al., 1997; Lecumberri et al., 2010). With a lower cognitive load, high proficiency speakers could have more cognitive resources available and, therefore, be more susceptible to priming than low proficiency L2 speakers (Berry & Ernestus, 2018).

In fact, the role of L2 proficiency on phonetic accommodation triggered by exposure has been documented in non-native contexts. For example, Berry and Ernestus (2018) investigated phonetic accommodation in Spanish learners of English towards Dutch learners of English. Spanish speakers with a higher level of proficiency level in English overall were found to accommodate at a greater rate to the interlocutor's pronunciation of the vowels under study than those with a lower proficiency level. In contrast, little is known about how proficiency may affect the degree of phonetic accommodation in interactions between a non-native speaker and a native speaker of the L2.

5.1.3.2. Does proficiency hinder accommodation triggered by feedback?

The presence of feedback during an interaction has also been argued to be a trigger for phonetic accommodation to take place. According to this view, accommodation is an audience design phenomenon (Bell, 1984): speakers accommodate their speech in order to ensure understanding from the interlocutor, and phonetic accommodation takes place when the communicative success is threatened (Branigan et al., 2010; Burnham et al. 2010; Gessinger et al. 2019). Feedback implying misunderstanding from the interlocutor can clearly signal to the speaker that the interlocutor is struggling to understand the message, a situation that can be overcome by phonetic adaptation.

In accordance with this view, in non-native contexts, the effect of feedback, either implicit or explicit, has been found to be an effective factor in L2 sound learning (Aliaga-García & Mora, 2008; Dlaska & Krekeler, 2013; Kartushina et al., 2015; Saito & Lyster, 2012). In L2 classroom settings, the presence of feedback has been found to make

L2 learners adapt their speech in the L2 to sound more native-like (Aliaga-García & Mora, 2008; Dlaska & Krekeler, 2013; Kartushina et al., 2015; Saito & Lyster, 2012).

In interactions involving non-native speakers, the effectiveness of feedback in triggering phonetic accommodation could depend on the level of the speaker's L2 proficiency. Lower proficiency speakers usually have a more accented pronunciation of the L2 sounds than higher proficiency speakers. As a consequence, because a more accented pronunciation may threaten the success of the conversation more, low proficiency speakers may be more motivated to accommodate to the native speaker in order to ensure the communicative success of the conversation.

5.1.4. Proficiency and the type of interlocutor

Both exposure and feedback can trigger phonetic adaptations in native and non-native speakers, but their effects may be modulated by the context in which the interaction takes place. Exposure has been argued to be more effective in triggering phonetic accommodation to human interlocutors than to computer interlocutors (e.g., Cohn et al, 2019; Chapter 3). Cohn et al. (2019) found that speakers accommodated more to human voices than to computer voices, even when they received the same amount of exposure from both sources. In contrast, feedback is more effective in triggering phonetic adaptation in human-computer interactions than human-human interactions (Burnham et al., 2010, Chapter 3). Speakers may be aware of the limitations of computers (which generally cannot adapt their perception to each human user they interact with). Therefore, a greater degree of phonetic accommodation in human-computer interactions could be the result of speakers ensuring understanding (Branigan et al., 2010).

As discussed in the previous sections, the role of proficiency on phonetic accommodation has been investigated before but, to our knowledge, no study has yet studied how the effect of proficiency on phonetic accommodation may be modulated by the type of interlocutor participating in the interaction. It is unclear whether the effect of proficiency on phonetic adaptation triggered by exposure and feedback may differ when non-native speakers interact with a human native speaker or with a computer.

5.1.5. Research questions

The present study investigates the role of proficiency in L2 sound learning through phonetic adaptation, triggered by exposure and feedback, in human-human and human-computer interactions. We address two research questions:

> RQ1: Does proficiency facilitate or hinder L2 learners' phonetic accommodation triggered by exposure to native pronunciation and/or feedback?
>
> RQ2: Does the type of interlocutor –human or computer– modulate the effect of proficiency on L2 phonetic accommodation?

Based on previous literature, we have two opposite hypotheses for RQ1. On the one hand, given that accommodation is assumed to be dependent on the cognitive resources available to the speaker (Costa et al., 2008), we expect high proficiency speakers to show more phonetic accommodation when receiving exposure to the native pronunciation than low proficiency speakers. On the other hand, given accommodation as an audience design phenomenon, we may expect low proficiency speakers to show more accommodation because they have more room to adapt their speech to native speech.

With respect to RQ2, based on Branigan et al. (2010), who argued that automatic processes, such as priming, are more effective in triggering phonetic accommodation in human-human interactions than in human-computer interactions, we expect the effect of proficiency in modulating the accommodation triggered by exposure to be stronger in human-human settings than in human-computer settings. Branigan et al. (2010) also argued that speakers tend to show more accommodation to ensure understanding in human-computer settings than in human-human settings. Therefore, we expect the effect of proficiency on the accommodation triggered by feedback on the learner's pronunciation to be stronger in the human-computer setting than in the human-human setting.

5.1.6. This study

In order to answer the research questions, we reanalyzed the data from an existing experiment designed to study phonetic accommodation in Dutch learners of English (Chapter 3). In this experiment, Dutch learners of English took part in a conversation with either a human or a computer interlocutor. In each of the settings –human-human or human-computer–, participants were randomly assigned to one of four conditions, combining the presence or absence of exposure to the native-like pronunciation of a specific L2 contrast –/æ, ɛ/–, and of feedback from the interlocutor implying that the learner's pronunciation of the contrast was not clear enough. The conditions thus represented a control condition (where participants did not receive exposure or feedback), an OnlyExposure condition (where they received only exposure but no feedback, an OnlyFeedback condition (where they received only feedback but no exposure) and a Combined condition (where they received both exposure and feedback). In Chapter 3, the learners' pronunciation of the contrast was analyzed before and after the interaction and accommodation was measured as adaptations made to the height (second formant) and frontness (first formant) of the vowels under study. The results from Chapter 3 showed differences in phonetic accommodation depending on the setting: in human-human settings, phonetic accommodation was triggered by exposure or feedback, in isolation from each other; in human-computer settings, in contrast, phonetic accommodation seemed to be triggered by feedback, both in isolation and in combination with exposure.

The current study differs from Chapter 3 in two aspects. First, we focused on the effect of the L2 proficiency level on phonetic accommodation in order to determine whether the effect of exposure and/or feedback on phonetic accommodation is modulated by how proficient the L2 learners are in English. Because we only have human judgments of the accentedness of some of the speakers (from Chapter 4), we estimated the proficiency level in the L2 with a series of assessments, including an objective measure, LexTale score (Lemhöfer & Broersma, 2012), and six subjective self-rated proficiencies in different language skills.

A second difference with the study presented in Chapter 3 is the way phonetic accommodation, affecting /æ, ɛ/, is assessed. Dutch speakers have been found to merge the vowels under study into an in-between vowel (Elsendoorn, 1985; Wang & van Heuven, 2006). Interestingly, these learners have also show asymmetrical knowledge of the height and frontness distinction of the two vowels: they can distinguish these two vowels in terms of height but not in terms of frontness (Troncoso-Ruiz et al., 2019). In Chapter 3, the focus was on whether participants could adapt the height and frontness dimensions differently. In the current study, phonetic accommodation is assessed by analyzing the distance between the two vowels instead because the focus is not on whether the two dimensions are adapted independently, but on whether the end result (i.e., enhancing the distinction between the two vowels) is affected by proficiency.

5.2. The dataset

This section describes the properties of the data set from Chapter 3 that are relevant for the present study, including the description of the participants, the tasks, and the stimuli involved in the tasks. For more details, see Chapter 3.

5.2.1. Participants

A sample of 96 female participants (M_{age}= 21.5 years, SD= 2.28) were recruited from the student population of Radboud University. None of the participants reported having any learning, hearing or reading disabilities. All of them were compensated financially for their participation in the study.

No participant was part of an English-taught program. An objective measure of their proficiency, the LexTale test (Lemhöfer & Broersma, 2012), indicated that, overall, participants' proficiency ranged from 48.75 to 97.50, which corresponds to a range from B1 and lower to C2 in the Common European Framework of reference for languages (Council of Europe, 2001). The distribution of the LexTale scores (see Table 1) does not differ statistically per experimental condition or per setting.

Table 1. Information about the LexTale score distribution (including the minimum LexTale score, the maximum LexTale score and the range of the LexTale scores) for the participants in the four conditions (Control, OnlyExposure, OnlyFeedback, Combined) of the two settings (human-human and human-computer)

Setting	Condition	Min. Lextale score	Max. Lextale score	Range of LexTale score
Human-computer	Control	60	88.8	28.8
Human-human	Control	58.8	83.8	25
Human-computer	OnlyExposure	48.8	91.2	42.5
Human-human	OnlyExposure	60	97.5	37.5
Human-computer	OnlyFeedback	68.8	90	21.2
Human-human	OnlyFeedback	55	90	35
Human-computer	Combined	63.8	88.8	25
Human-human	Combined	65	91.2	26.2

5.2.2. Materials and Procedures

The overall structure of the experiment consisted of a pre-test sentence completion task, a tone discrimination task, the interaction stage (where participants engaged in the interaction with a confederate, human or computer), a post-test sentence completion task, the LexTale test, and a battery of questionnaires about the participants' demographics and linguistic background. The tone discrimination task is not relevant for the present study and will therefore not be further discussed (see Chapter 3 for a full description). We will base our acoustic analyses on the data from the pre- and post-task.

5.2.2.1. Sentence Completion Tasks

The participants performed the sentence completion task immediately before (pre-test) and after (post-test) the interaction. Participants were told that, due to the length of the task, it was split into two parts. Participants were presented with 48 incomplete sentences on a computer. Each sentence was presented in isolation in the middle of a laptop screen (Calibri, pt. 24) on PowerPoint presentation. Participants were told to, first, read the sentence in silence, find a suitable ending for the sentence and, then, read the sentence together with their suitable ending out loud. The speech of the participants was recorded with a Seinheisser K-6 microphone, at a sampling rate of 44.1 kHz. This task was self-paced: participants could proceed to the next sentence by pressing the space bar on the keyboard. On average each part of the task, pre-test and post-test, took 10-15 minutes.

In the incomplete sentences (sentence beginnings) used for these tasks, a selection of 24 English minimal word pairs of /æ, ɛ/ were embedded (see Appendix 1 for the full list of words). The mean zipf value of the 48 selected words, which indicates the frequencies of the words in a 1-7 scale, was 4.81 (SD=.661) according to van Heuven et al. (2014). All of the words from this list of minimal pairs were monosyllabic words, except for "cattle" and "kettle". Furthermore, words representing four additional English contrasts, including other sounds, were selected. These contrasts included three consonant contrasts /m, n/, /p, b/ and /r, l/, and another vowel contrast /ɪ, i:/. These four filler contrasts were are assumed to be easier to distinguish for Dutch learners of English than the /æ, ɛ/ contrast. The words selected for each of these contrasts also consisted of 24 minimal pairs

Two different sentence beginnings were created for each of the 48 critical words. The two beginnings containing the same critical word were designed to be as similar as possible in terms of syntactic structure, to make the context in which the critical token was produced as comparable as possible. Overall, the 96 sentence beginnings varied in terms of length and difficulty to complete. One of the two sentence beginnings was used in the pre-test and the other sentence beginning in the post-test task in order to ensure that participants produced the same vowels in the same context before and after the interaction, allowing for the assessment of any phonetic adaptation. A total of 12 different lists were used to distribute the 96 sentence beginnings into the pre-test and post-test tasks, ensuring that all of the sentence beginnings appeared either in the pre-test or the post-test of the experiment. These lists were used in each of the conditions, and each participant was exposed to only one of the lists. The filler minimal pair words were also embedded in these experimental lists. Thus, in case participants noticed the repetition of words from pre-test to post-test, the presence of other minimal pairs would make it less evident for participants that the /æ, ɛ/ contrast was the critical focus of the experiment.

5.2.2.2. Interaction

During the interaction, the participant played a puzzle-solving game together with either a human confederate or a computer. This game required the participant and the confederate to share information on their screen to solve a series of puzzles. In the human-human interaction, the Ventriloquist paradigm (Chapter 2) was used in order to implement the same pre-recorded speech as in the human-computer interaction, without participants noticing. The participant and the human confederate were sitting in the same room across each other, making eye contact, only separated by a table between them. In order to maintain the illusion that the pre-recorded speech was live, the participants were equipped with headphones and microphones and were told that they would hear the interlocutor through their headphones. In reality, whenever the human confederate needed to speak, she would lean towards the microphone, placed at the edge of the table, moving her face behind a monitor and away from the sight of the participant, and subtly press a key to play one of the audio recordings. The audio would then be played through the participant's headphones.

In the computer-human setting, participants performed the same task but with a computer instead of a human confederate. Participants sat in a smaller recording booth by themselves and were also equipped with a microphone and headphones. They were told they would be interacting with a smart computer while playing a game. In reality, the same confederate who played the human interlocutor in the human-human setting was in control of the responses of the computer. Sitting outside of the booth, she could trigger the playing of the same pre-recorded audio samples to the participants as in the human-human setting.

5.2.2.2.1. The pre-recorded speech used

The pre-recorded speech played in both settings was carefully scripted to avoid the presence of the critical contrast –/æ, ɛ/– and it consisted of two categories of audios samples: trial-related audios and spontaneous audios. The trial-related audios were fragments of speech specifically designed for each trial of the puzzle-solving game; so, they were specific to the information displayed on the screen. The spontaneous audios were used to maintain the illusion of a spontaneous conversation and designed to help the confederate respond to any spontaneous contributions to the conversation by the participant. The spontaneous speech was divided into nine categories: affirmative answers (“Yup”), negative answers (“No”), answers indicating the confederate did not know (“I’m not sure”), requests for clarification, reassuring answers (“no worries”), utterances asking for repetition (“Could you repeat it?”), utterances reminding about the rules (“I don’t think you’re allowed to do so”) and utterances showing surprise for a new trial appearing on the screen. Each category was linked to a different key of the numeric keyboard (see see Chapter 2 for a full description of the audios used). The utterances were recorded by a 25-year-old female native speaker of English.

5.2.2.2.2. The puzzle-solving game

The game consisted of two types of trials: the Semantic Relation trials, designed to manipulate the exposure to the native-like pronunciation, and the Code Breaker trials, designed to implement the presence versus absence of feedback on the learner’s pronunciation of the contrast. In total, the game consisted of 69 trials: 13 Semantic Relation trials and 56 Code Breaker trials. The session started with 4 Code Breaker and 1 Semantic Relation trials as training trials to familiarize participants with the dynamics of the game.

5.2.2.2.3. The implementation of Exposure

Because of the scripted nature of the pre-recorded speech used during the interaction, participants were never exposed to/æ, ɛ/ as produced by the interlocutor. The only part of the interaction where participants could receive exposure to the native-like pronunciation of this contrast was in the 12 Semantic Relation trials. In these trials, the screen of the participant displayed a picture, for instance including Barack Obama and Queen Elizabeth II of England. The screen of the confederate displayed a set of four words (e.g., "pin", "bin", "crown", or "clown"). In order to solve this type of puzzles, the participant and the interlocutor needed to share the information on their screens and discover which of the words displayed on the confederate's screen was semantically related to the picture in the participant's screen. Thus, during a trial, participants would start describing their picture and the confederate would play a pre-recorded audio indicating which words were displayed on her screen and which one she thought was related to the description provided by the participant.

In the conditions with exposure, the words the confederate read out loud included one of the 24 critical minimal pairs and one of the pairs from the filler minimal pairs, described in section 5.2.2.1, in all 12 trials (for example: "the words on my screen are bad, bed, crown or clown, so I think the right word is crown"*;* where "bad" and "bed" would be the critical words, and "crown" and "clown" would be the filler items). The critical words were always read in first and second position and they were never the words related to the picture displayed in the participant's screen. Assuming that participants could stop paying attention once they heard the word that matched their picture, the critical words occurred before the right answer, which ensured participants closely listened to the words. The order in which the critical words appeared (i.e., "bad"-"bed" or "bed"-"bad") was counter-balanced throughout the experiment.

In conditions without exposure, the four words displayed on the confederate's screen in the 12 Semantic Relation trials were filler minimal pairs. Thus, the pre-recorded audio that the confederate played in these trials did not contain any tokens of the critical vowels and participants were never exposed to the native-like pronunciation of the critical contrast during the interaction.

5.2.2.2.4. The implementation of Feedback

Feedback was implemented as corrective visual feedback on the participants' pronunciation during the CodeBreaker trials. In these trials, the participant saw four different geometric shapes on their screen. Each shape was linked to one word, located immediately below the shape. On the confederate screen, a sequence of shapes was displayed on the top, and a set of four words was displayed at the bottom. The participant and the confederate needed to share the information displayed on their screens to discover which of the shapes in the participant's screen could complete the pattern in the sequence of shapes displayed in the confederate's top part of the screen. Once they agreed on a shape, the participant would tell the confederate which word was linked to the selected shape, so that the confederate could select that word to proceed to the next trial. In 12 out of the 64 trials, the words displayed on the participant's and the confederate's screen consisted of a critical minimal pair word and a filler minimal word pair (described in section 5.2.2.1), for instance "bad", "bed", "clown", "crown". In these 12 experimental trials, the word linked to the right shape was always the token containing the /æ/, for example "bad".

In conditions with feedback, participants could see a grey rectangle on their screen appearing around the word selected by the confederate. The confederate always selected the word indicated by the participant, except in the 12 experimental trials, where the target word belonged to the critical minimal pair words. In these experimental trials, participants had to instruct the confederate to click on the /æ/ word, but the confederate systematically clicked on the word containing the /ɛ/ vowel. This visual feedback was designed to make participants aware that their pronunciation of the /æ/ vowel was not clearly distinct from the /ɛ/ vowel to a native speaker of English. In order to make the interaction as similar as possible for all participants, the feedback remained negative (the confederate systematically selected the /ɛ/ word instead of the /æ/ word) in all 12 trials.

In conditions without feedback, no grey rectangle was displayed in any of the in any of the 64 CodeBreaker trials. Thus, participants were never made aware of the fact that their pronunciation could be misunderstood by a native speaker of English.

5.3. Measuring proficiency and accommodation

5.3.1. Proficiency assessment

The two opposite hypotheses of the effect of proficiency on non-native accommodation are based on slightly different aspects of proficiency. On the one hand, there is the hypothesis assuming that a higher proficiency leads to more accommodation because a higher proficiency implies the availability of more cognitive resources. This hypothesis concerns a general proficiency, including the ease of lexical access. On the other hand, there is the hypothesis concerning the nativelikeness of the pronunciation, assuming that a less native-like pronunciation leads to more accommodation. Preferably, the proficiency measure that we use to analyze the data captures both aspects of proficiency (availability of cognitive resources and the nativelikeness of the pronunciation).

We base ourselves on a series of proficiency measures. We do so by combining the measures that we collected during the experiment, which not only yields a good measure of general proficiency but also correlates best with human evaluations of nativelikeness for some of the speakers (see below), which we collected in Chapter 4. There, the nativelikeness ratings were only collected for the subset of the data corresponding to the human-human setting. We based the proficiency measure in this study on the LexTale score and six self-rated proficiency measures collected via a questionnaire.

The LexTale test (Lemhöfer & Broersma, 2012) is a relatively short and easy task, which has been found to correlate with standardized classifications of proficiency, such as the Common European Framework of reference for languages (Council of Europe, 2001). At the end of the experimental sessions, participants were asked to complete the English version of the LexTale task on a laptop online. In this task, participants see 60 words, including words and non-words, displayed on a laptop screen. They have to indicate whether the words are real English words by clicking the words “yes” or “no” displayed on the screen using a mouse. This test took approximately 5 minutes. The LexTale score, which is based on the learners' vocabulary size in the L2, could arguably be biased towards the written component of the language, i.e., reading and writing skills.

The questionnaire that the participants completed contained questions about their L1 and L2 experience. There were four questions where participants used a 5-point scale to rate their own English skills, namely, writing, listening, speaking and reading. Furthermore, participants were asked to rate their own accent in English, namely the strength of their non-native accent and how often they were identified as non-native speakers by native speakers of the L2. Self-evaluations have disadvantages. Most importantly, assessing language skills in an objective way is usually a difficult task. Non-native speakers' ratings of their own proficiency in an L2 have been found to be acceptably accurate but dependent on other factors, such as the speakers' L1 (Diamond et al., 2019), and little is known about how well self-reported proficiency reflects the learners' L2 pronunciation.

Principal Component Analysis (PCA) was used to compare the different proficiency measures and to derive one single proficiency measure. PCA was computed using the "prcomp" function from the R built-in "stats" package (R Core Team, 2020) (version 3.5.0). Table 2 shows the percentage of variance explained by the components and the contribution of the seven proficiency measures to each of the principal components derived from the analysis. PC1, the component with the highest variance of data explained, is the component that most equally represents all the different measures of proficiency used.

Table 2. Contribution (in percentages) of each of the proficiency measures to each of the components and the percentage of variance explained by each of the components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Percentage of variance explained	*53.38*	*17.48*	*9.10*	*6.81*	*5.79*	*4.67*	*2.77*
Lextale	9.00	22.09	50.78	9.16	3.87	0.52	4.58
ProficiencyListening	16.22	9.83	3.61	6.88	42.80	3.42	17.23
ProficiencyWriting	16.38	0.74	29.05	18.30	13.00	7.57	14.96
ProficiencyReading	15.23	19.14	4.24	11.06	2.72	10.23	37.37
ProficiencySpeaking	19.03	2.92	0.02	0.50	5.81	66.06	5.66
AccentStrength	13.34	20.03	3.85	15.65	26.36	7.58	13.18
IdentificationAsNon-Native	10.79	25.24	8.45	38.45	5.43	4.62	7.02

In order to understand how representative the measures collected as well as the PCs were of the participants' nativelikeness, they were compared to native speakers' judgements of the participants in the human-human setting with respect to their pronunciation (collected in Chapter 4; no judgments are available for the participants in the human-computer task). These judgements were provided by a total of 107 native speakers of English, who listened to the pronunciation of the experimental words with /æ, ɛ/ as produced by the Dutch learners of English in the human-human setting of the experiment. While hearing each word in isolation, the native speakers saw the orthographic transcription of the word intended by the Dutch learner on the screen. They rated the nativelikeness of the pronunciation on a 5-point scale, using a keyboard. An average score was calculated for each non-native speaker.

Table 3 shows the correlation of the seven principal components with the nativelikeness scores provided by the native speakers. This table shows that the PC1 component is the assessment with the highest correlation with the native judgements. PC1 not only represents a general measure of proficiency as it incorporates all measures collected but also is the component that best associates with the nativelikeness. Thus, for this study, PC1 was selected as the one measure of proficiency.

Table 3. Table showing the correlation between the PCs and the native-like judgements.

Principal Component	Correlation with native speakers' judgments
PC1	-0.28
PC2	0.19
PC3	0.07
PC4	-0.12
PC5	0.06
PC6	0.03
PC7	0.01

5.3.2. Measuring phonetic accommodation

Since Dutch learners tend to merge the English /æ/ and /ɛ/ vowels into an in-between sound, phonetic adaptation to a native speaker could be interpreted as the degree to which they unmerge this contrast in their pronunciation. Thus, the degree of accommodation was measured as the distance between the pronunciation of these two vowels, /æ, ɛ/. Greater distance in the post-test than in the pre-test was interpreted as evidence for accommodation.

A total of 4,130 word tokens containing the /æ, ɛ/ contrast were extracted from the participants' recordings in the pre-test (2,065) and post-test (2,065) sentence completion tasks. Only experimental minimal pair words (described in section 5.2.2.1) were selected, e.g., "bad" and "bed", to ensure that the context in which the two vowels compared were produced was as similar as possible. For each participant, a mean of 21.51 words (sd=0.78) were extracted from both the pre-test and post-test. The first (F1) and second (F2) formants of their vowels were extracted at the midpoints with a script on Praat (Boersma & Weenink, 2019). Then, using the F1 and F2 values, the Euclidean distance between the minimal word pairs ("bad" and "bed") for each participant was calculated using the "eucl_dis()" function from the "joeyr" package (version 0.1.3) (Stanley, 2018) in R (R Core Team, 2020)(version 3.5.0).

5.4. Results

5.4.1. Descriptive data

Figure 5 shows the pretest-posttest difference in vowel distance by the proficiency level, measured with PC1, for learners in the four different conditions (Control, OnlyExposure, OnlyFeedback, Combined) in the two settings (human-human and human-computer).

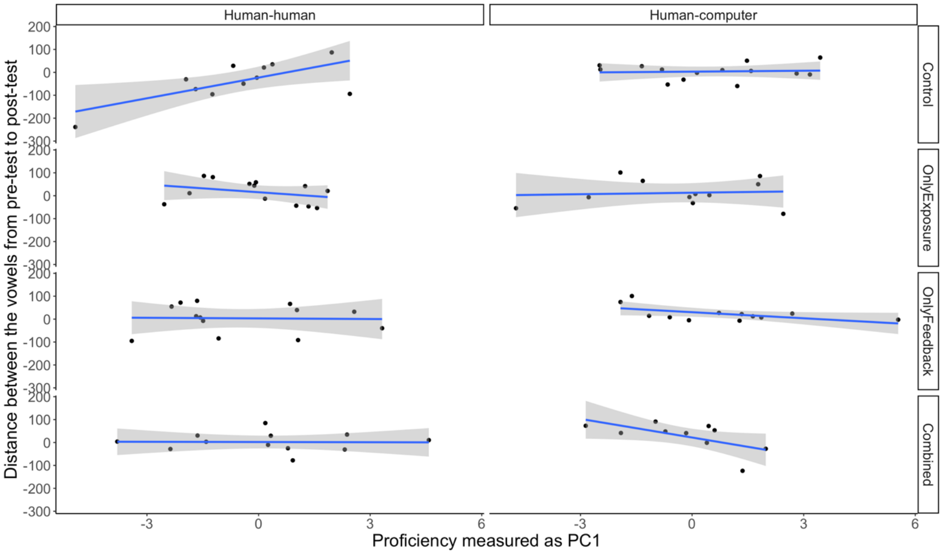


Figure 5. Scatterplot where each dot represents the average mean per participant (in each condition and setting) of the difference from pre-test to post-test of the distance between the two critical vowels, based on the F1 and F2 values.

In the Control condition of the human-human setting, L2 proficiency seems to foster an increase in the difference between vowel distance from the pre-test to the post-test: the vowels become more different the higher a participant's proficiency score. In the rest of the conditions (OnlyExposure, OnlyFeedback and Combined condition), the difference from pre-test to post-test does not seem to be modulated by the proficiency of the L2 learners.

In the human-computer setting, proficiency does not seem to play a role in the Control, OnlyExposure and OnlyFeedback conditions. In contrast, in the Combined condition, proficiency level seemed to modulate the pretest-posttest difference in distance between the vowels: participants seemed to decrease the distance between the two vowels in the post-test as compared to the pre-test more the higher their proficiency level. This effect of proficiency seems to be opposite to the tendency found in the Control condition of the human-human setting.

5.4.2. Statistical analysis

In order to investigate whether the tendencies found in Figure 5 are statistically significant, a linear mixed effect model was fitted to the data using the "lmer" function from the "lme4" package (Bates et al., 2015) (version 1.1.21) in R (R Core Team, 2020) (version 3.5.0). In particular, the model was designed to investigate whether the proficiency level had an effect on the pretest-posttest difference between the vowel distance for a given word pair by a given participant and whether this effect varied depending on the condition and setting to which participants had been assigned. Thus, the dependent variable was the distance between the two vowels for a given word pair by a given participant. The independent variables were Test (pre-test, post-test), Proficiency (PC1), Condition (Control, OnlyExposure, OnlyFeedback, Combined) and Setting (Human-human, Human-computer). Random intercepts were included for Speaker and Word. A random slope for Test by Speaker was included in order to prevent any significant effects being driven by individual participants. Similarly, in order to prevent any effects being reliant only on individual words, a random slope for Test by Word was also included. The final model was refitted removing the outliers of the model (2.5 standard deviations) to confirm the significant effects and interactions were not relying on outliers.

Table 4. Simplified output (with only the interactions relevant for the RQs) of the regression model fitted to the vowel distance data (for full version see Appendix 28). Values in bold are considered significant.

	Estimate	*Std. Error*	*df*	*t value*	*P-value*
(Intercept)	**215.32**	**20.21**	**90.67**	**10.65**	**<0.001**
Test(Post)	-22.63	16.81	86.17	-1.35	0.182
PC1	**-37.10**	**9.77**	**79.72**	**-3.80**	**<0.001**
Test(Post):PC1	**30.45**	**8.14**	**78.85**	**3.74**	**<0.001**
Test(Post):PC1:Condition(OnlyExposure)	**-41.60**	**13.37**	**80.93**	**-3.11**	**0.003**
Test(Post):PC1:Condition(Combined)	**-30.78**	**10.58**	**78.99**	**-2.91**	**0.005**
Test(Post):PC1:Condition(OnlyFeedback)	**-31.40**	**10.92**	**79.40**	**-2.87**	**0.005**
Test(Post):PC1:Setting(Computer)	**-29.18**	**10.97**	**78.69**	**-2.66**	**0.009**
Test(Post):PC1:Condition(OnlyExposure): Setting(Computer)	**42.56**	**17.00**	**79.75**	**2.50**	**0.014**
Test(Post):PC1:Condition(Combined): Setting(Computer)	2.32	17.33	78.22	0.13	0.894
Test(Post):PC1:Condition(OnlyFeedback): Setting(Computer)	21.30	15.08	78.63	1.41	0.162

Table 4 presents part of the output of the model with the pre-test distance between the two vowels and the Control condition of the human-human setting on the intercept. This model revealed significant interactions between Test, PC1, Condition, and the setting, indicating that the effect of proficiency differed depending on the condition and on whether participants interacted with a human or a computer interlocutor. To ease the interpretation of results, we will first discuss the findings related to the human-human setting and then the findings related to the human-computer setting.

Human-human setting

Table 4 shows a significant interaction between Test and PC1. This interaction indicates that, in the condition and setting at the intercept (the Control condition in the human-human setting), the difference from pre-test to post-test in vowel distance was modulated by the proficiency level of the L2 speakers. In particular, participants with a higher proficiency level increased the difference between the two vowels more than those with a lower proficiency. The model also shows significant interactions between Test, PC1 and the other three conditions for the human-human setting, showing that the effect of PC1 on Test was different in the other conditions. In order to interpret these interactions, the model was relevelled to include each of the conditions on the intercept (see Appendixes 29, 30 and 31) The relevelled models revealed no significant interactions between Test and PC1, meaning that no evidence was found for the level of proficiency to modulate the pretest-posttest difference in the OnlyExposure, OnlyFeedback and Combined conditions.

Human-computer setting

In order to interpret how the effect of proficiency on phonetic accommodation was different in the human-computer setting, the original model (presented in Table 4) was relevelled to include the human-computer setting on the intercept (Table 5).

Table 5 includes part of the output of the model with the Control condition of the human-computer setting on the intercept. This model does not reveal a significant interaction between Test and PC1, and there is thus no evidence that proficiency modulates phonetic accommodation in the Control Condition. The lack of significant interactions between Test, PC1 and the OnlyExposure and between Test, PC1 and the OnlyFeedback conditions suggests that proficiency did not modulate the pretest-posttest difference in the participants of these conditions in the human-computer setting either. In contrast, the model did reveal a significant interaction between Test, PC1 and the Combined Condition, implying that in this condition the relation between Test and PC1 was different from that in the Control condition. In order to interpret this interaction, the model was relevelled to include the Combined condition of the human-computer setting on the intercept. The relevelled model (see Appendix 33) revealed a statistically significant interaction between Test and PC1 ($\beta = -27.191$, $p = 0.022$), indicating that in this condition the pretest-posttest difference was modulated by the proficiency of the L2 learners. In particular, the difference between pre-test and post-test vowel distance was smaller when the proficiency level was higher. A higher proficiency thus hindered phonetic adaptation.

Table 5. Simplified output (including only the interactions relevant for the RQs) of the regression model fitted to the vowel distance data with the Control condition of the human-computer setting on the intercept (see Appendix 32 for full model).

	Estimate	*Std. Error*	*Df*	*t value*	*Pr(>|t|)*
(Intercept)	**147.45**	**17.93**	**91.35**	**8.221**	**<0.001**
Test(Post)	3.29	14.84	83.31	0.222	0.825
PC1	-2.50	8.83	79.60	-0.283	0.778
Test(Post):PC1	1.27	7.35	78.51	0.172	0.864
Test(Post): PC1:Condition(Combined)	**-28.46**	**13.73**	**77.77**	**-2.073**	**0.041**
Test(Post): PC1:Condition(OnlyFeedback)	-10.10	10.40	77.80	-0.971	0.334
Test(Post): PC1:Condition(OnlyExposure)	0.96	10.49	77.90	0.092	0.927
Test(Post):PC1:Setting(Human)	**29.18**	**10.97**	**78.69**	**2.661**	**0.009**
Test(Post):Condition(Combined): Setting(Human)	6.97	30.98	80.19	0.225	0.823
Test(Post):Condition(OnlyFeedback): Setting(Hum.)	-2.04	30.63	79.76	-0.067	0.947
Test(Post):Condition(OnlyExposure): Setting(Hum.)	28.38	30.34	80.25	0.935	0.352
PC1:Condition(Combined):Setting(Human)	9.66	20.83	79.50	0.464	0.644
PC1:Condition(OnlyFeedback): Setting(Human)	**45.98**	**18.11**	**79.65**	**2.539**	**0.013**
PC1:Condition(OnlyExposure): Setting(Human)	**48.48**	**20.36**	**80.06**	**2.381**	**0.02**
Test(Post):PC1:Condition(Combined): Setting(Hum.)	-2.32	17.33	78.22	-0.134	0.894
Test(Post):PC1:Condition(OnlyFeedback): Setting(Human)	-21.30	15.08	78.63	-1.412	0.162
Test(Post):PC1:Condition(OnlyExposure): Setting(Human)	**-42.56**	**17.00**	**79.75**	**-2.504**	**0.014**

In summary, proficiency was found to modulate the phonetic adaptation shown by non-native speakers. The modulation depended on whether participants interacted with a human confederate or a computer interlocutor. In the human-human computer, L2 proficiency fostered learning but only when participants where in the control condition and did not receive exposure to the native-like pronunciation or feedback on their pronunciation of the contrast. In the computer-setting, proficiency had the opposite effect, hindering phonetic adaptation but only when they received both exposure and feedback.

5.5. Discussion

The present study aimed to study the mechanisms of priming and audience design underlying phonetic accommodation in non-native speakers. We investigated the effect of L2 proficiency on the degree of phonetic adaptation displayed by Dutch learners of English as the result of receiving exposure to the native pronunciation of a contrast, /æ, ɛ/, and feedback on their own pronunciation of the same contrast. The effect of L2 proficiency on phonetic accommodation was compared in human-human and human-computer interactions. Here, we used a combinatory measure of proficiency based on Principal Component Analysis to assess the L2 proficiency of the participants. This measure well reflects general proficiency and also showed the highest correlation with human evaluations that we have for some participants of the nativelikeness of their pronunciation.

5.5.1. Does proficiency facilitate or hinder L2 learners' phonetic accommodation triggered by exposure to native pronunciation and/or feedback?

The findings of this study indicate that, in the human-human setting, proficiency modulated the degree of phonetic adaptation shown by participants in the Control condition. In this condition, participants did not receive exposure to the nativelike pronunciation of the critical contrast, /æ, ɛ/, or negative feedback on their pronunciation of the same contrast. Still, high proficiency speakers were found to adapt their pronunciation of the vowels, enlarging the distance between the two sounds in the post-test, more than low proficiency speakers. These findings, however, cannot directly give evidence for the accommodation as an automatic mechanism perspective (e.g., Costa et al., 2008; Pickering & Garrod, 2004;), since the adaptation in this case was not triggered by direct exposure to the native-like pronunciation of the vowels under study.

One possible explanation for the presence of adaptation modulated by proficiency in the Control condition could be related to the nature of the task. While playing the interactive game, participants had to instruct the confederate to click on a specific word. In twelve of the trials, the participants were visually exposed to /æ, ɛ/ minimal pairs on their screen and they had to instruct the confederate to click on the word containing the /æ/ sound (for example, "bad") in a way that was distinctive enough from the other words displayed on the screen, including the /ɛ/ counterpart, e.g., "bed". Being visually exposed to the minimal pairs could have raised participants awareness of the difficulty of producing these two vowels distinctively.

The effect of proficiency indicates that the awareness-raising element of the task, i.e., noticing that the pronunciation of the two critical sounds is difficult, was processed differently by high proficiency and low proficiency speakers. In particular, high proficiency speakers adapted their speech more than low proficiency speakers in the control condition. It could be the case that the higher amount of cognitive resources available led to a higher level of awareness of the difficulty of the contrast in high proficiency speakers than in low proficiency speakers. The higher level of awareness then could have triggered more adaptation in participants with a higher proficiency level. Alternatively, both high and low proficiency speakers could have become equally aware of the difficulty of the contrast during the interaction, but the higher amount of cognitive resources allowed high proficiency speakers to react to the awareness of the difficulty of the contrast more than low proficiency speakers. Thus, although the adaptations displayed in this condition could not be the result of priming –due to the lack of exposure–, it seems that proficiency facilitated phonetic adaptation triggered by awareness.

In the condition where participants were exposed to the /æ, ɛ/ contrast, there was no effect of proficiency on accommodation. Previous literature has argued that speakers with more cognitive resources available are more susceptible to priming (e.g., Berry & Ernestus, 2018; Costa et al., 2008). This idea is thus not supported by our study as no evidence was found for high proficiency speakers, who have a cognitive advantage over low proficiency speakers, being more susceptible to priming than low proficiency speakers.

When participants received feedback on their pronunciation, we did not find an effect of proficiency on accommodation either. The audience design perspective on phonetic accommodation assumes that phonetic accommodation is driven by the communicative success (Bell, 1984). Thus, speakers with a bigger pronunciation gap with the interlocutor, i.e., a more accented pronunciation, would accommodate more because the pronunciation gap would pose a bigger threat to the communicative success. This hypothesis is thus not supported by our findings in the human-human setting: the results of this study do not show differences in the participants' accommodating behavior depending on their proficiency level.

In this study, we also included a condition, the Combined condition, where participants received both exposure and feedback during the interaction. The results showed no differences in accommodating behavior depending on the proficiency level of the non-native speakers. Thus, no evidence was found for the cognitive resources available to the speakers or the pronunciation gap between them and the human confederate to modulate how participants processed the combination of exposure and feedback during the interaction.

Overall, the results of this study seem to suggest that, regardless of their proficiency level, participants interacting with the human interlocutor processed exposure to the native-like pronunciation and/or feedback on their pronunciation similarly. Thus, no evidence was found for phonetic accommodation to be driven by the amount of cognitive resources available or by the phonetic distance between the non-native speaker and the interlocutor in those three conditions (OnlyExposure, OnlyFeedback and Combined condition). One possible explanation as to why the level of proficiency was not found to modulate the accommodation triggered by exposure and/or feedback but was a relevant factor in the Control condition could be related to the difficulty of the task in the different conditions. In the Control condition, participants did not receive and therefore did not have to process exposure to the critical sounds and they were not made aware of the accentedness of their pronunciation through feedback. Although participants may have become aware of the difficulty of the contrast by being visually exposed to the minimal pairs (as previously discussed), the conversation in this condition could be considered to be less demanding than in the other conditions. It could be the case that receiving exposure and/or feedback was equally challenging for participants in the experiment, regardless of their proficiency level. Therefore, proficiency only made a difference when the participants' cognitive resources were not overwhelmed, in the Control condition.

Alternatively, it could be the case that the range of proficiency levels in this setting was not wide enough in order for proficiency to make a difference. As described in Table 1 of section 2.1, the proficiency levels of the participants varied from B1 to C2 in the CEFR. Hence, the low proficiency end of the proficiency spectrum was set in an intermediate step, B1. The high end of the spectrum, in contrast, was set in the most advanced category for language learners, C2. It could be the case that, in order for proficiency to significantly modulate adaptation triggered by exposure and feedback in human-human settings, the endpoints of the proficiency range need to be further apart from each other. For example, this study could have included Dutch speakers with and A1 level of English.

5.5.2. Does the type of interlocutor –human or computer– modulate the effect of proficiency on L2 phonetic accommodation?

The analysis reported in section 5.4.2 showed statistically significant differences between the human-human and the human-computer settings, following previous literature investigating how phonetic accommodation differs in these two settings (Branigan et al., 2010). In the human-computer setting, unlike in the human-human setting, proficiency did not seem to have an effect on the degree of phonetic adaptation displayed by the non-native speakers in the Control condition. Similarly, unlike previous literature (e.g., Berry & Ernestus, 2018), we did not find significant differences depending on the level of proficiency in the accommodation triggered by exposure to the interlocutor in the human-computer setting. Therefore, no evidence was found in this setting for accommodation to be an automatic mechanism modulated by the cognitive resources available (e.g., Pickering and Garrod, 2004; Costa et al., 2008). The effect of feedback in isolation on phonetic adaptation was not found to be modulated by proficiency in the human-computer setting either. No evidence was found in this condition showing that low proficiency speakers, with a greater distance in terms of pronunciation from the native-like pronunciation, accommodated more than high proficiency.

In contrast, phonetic accommodation to the computer interlocutor seemed to be modulated by proficiency when the participants received feedback on their own pronunciation together with exposure to the native-like pronunciation of the critical contrast. The fact that the combination of feedback and exposure triggered more phonetic adaptation in low proficiency speakers than in high proficiency speakers indicate that, in contrast with Costa et al. (2008), the cognitive advantage in high proficiency speakers did not make them more susceptible to be primed by exposure and, therefore, to show more accommodation. In fact, this finding provides evidence for the perspective describing non-native phonetic accommodation as an audience-design phenomenon (Bell, 1984). In this condition, low proficiency speakers, whose pronunciation may have posed a bigger threat to the communicative success, were found to accommodate more than high proficiency speakers.

One possible explanation for this finding could be that the presence of feedback and exposure together made participants both in the high and low end of the proficiency spectrum focus on the exposure equally. Since the feedback was always corrective, participants may have tried to search for clues on how to enlarge the distinction of the two vowels in the speech of the computer. Given that low proficiency speakers in general showed a greater distance to the interlocutor, they, then, adapted their speech more than high proficiency speakers to ensure the communicative success.

Alternatively, the combination of exposure and feedback could have been interpreted differently by high and low proficiency participants. For high proficiency speakers, whose pronunciation was almost native-like, the presence of consistent corrective feedback may initially have alerted them of the possible pronunciation gaps between them and the computer. However, the exposure to the native-like pronunciation, similar to their own could have reassured them about their pronunciation and made them realize that the feedback was not dependent on their actual production of the critical sound. Noticing the consistency could have made the experience of interacting with the computer interlocutor frustrating. Frustration is indeed one of the emotions that can hinder L2 learning (Hashemi, 2011; Huber & Arthur, 2016). The fact that the feedback was maintained corrective during the interaction may have gone unnoticed by low proficiency speakers, who knew that they had a more accented pronunciation.

The comparison between these two settings was based on identical interactions. Pre-recorded speech was used in order to ensure the input participants received, either from a human or a computer, was exactly the same. However, there were of course other aspects differing in the two settings that could have played a role and that may explain the differential effect of proficiency, rather than communicative success. In the human-human setting, participants were exposed to a human confederate sitting in the same room across a table. Thus, participants had more information to process in this context, such as facial expressions, gestures or the physical appearance of the confederate. Participants in the human-computer setting, in contrast, were sitting in a relatively smaller booth using a computer. Therefore, they did not have to process this extra information. As a consequence, participants in the Combined condition of the human-computer setting could have enjoyed more cognitive resources available, since they did not have to interact with a human in the same room, to process the presence of exposure and feedback combined and, as a consequence, could adjust their pronunciation.

The conclusions from this study highlight the importance of using ecologically valid settings resembling human-human interactions when studying the mechanisms underlying the phenomenon of phonetic accommodation. Previous studies investigating phonetic accommodation have made use of computer-based settings, assuming that the behavior observed in those experiments is reflective of what humans do when interacting with each other (e.g., Babel, 2012; Cohn et al., 2019, among others). The current study shows that phonetic accommodation, either triggered by exposure or feedback, is heavily dependent on the methodology used to recreate an interaction. Thus, this study calls for more ecologically valid settings to be used when studying human-human interactions.

5.5.3. Proficiency assessment

The current study investigated two different hypotheses about the effect of proficiency on non-native accommodation. The first hypothesis of accommodation as an automatic mechanism predicts high proficiency speakers to accommodate more than low proficiency speakers because higher proficiency is assumed to be translated into more cognitive resources available to the learner. The second hypothesis of accommodation as an audience design predicts that low proficiency speakers accommodate more than high proficiency speakers because their pronunciation gap with the native interlocutor would lead to a bigger threat to the communicative success. In order to test these hypotheses, the current study aimed to find a proficiency measure that reflected both the overall proficiency of the participants in the L2, linked to the cognitive resources available, and also the nativelikeness of their pronunciation. For this purpose, we used a measure combing the LexTale score, self-rated listening, speaking, writing and reading skills, self-rated pronunciation accentedness and self-rated identification as a non-native speaker (PC1).

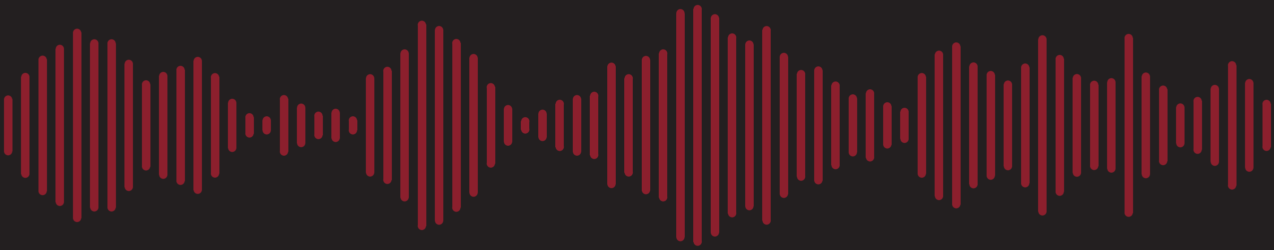
PC1 was found to be an acceptable measure for the purpose of this study. On the one hand, this component represented an overall proficiency measure, based on an objective assessment of learner's the vocabulary knowledge, LexTale, and the self-rating of other language skills, such as reading and writing. The general aspect of this measure could be argued to be indicative of the cognitive resources available. On the other hand, this component was also found to be associated with the nativelikeness of the learners' pronunciation as rated by native speakers of English. Thus, PC1 used in this study was both representative of the overall proficiency and of the nativelikeness of the pronunciation.

The results of the comparison between the principal component measuring proficiency and the nativelikeness judgements have two implications. First, these findings indicate that a component combining different proficiency measures can be a moderately reliably source to obtain proficiency ratings reflecting the nativelikeness of their pronunciation. Second, the combination of all proficiency measures in PC1 was found to show a higher correlation to the nativelikeness ratings than the other principal components, which were mapped on smaller subsets of the data. Thus, in order to assess the nativelikeness of L2 learners, it seems that the combination of different subjective measures and LexTale is a better substitute for the nativelikeness of the proficiency than the other measures.

The conclusions of this study, however, also have some limitations. First, the proficiency measure used, PC1, is assumed to reflect on the cognitive abilities of the learners because it is based on the overall proficiency. However, in this study, this connection has not been tested directly. In order to verify this connection, a measurement of resources available in the L2 could have been used to analyze the data. Since cognitive resources cannot be directly measured, proxies of the cognitive abilities could have been used for this analysis. For example, the design of the experiment could have included an attention task, memory task or inhibitory control task (Darcy et al., 2014) to measure the L2 speakers' cognitive abilities. Similarly, the data from the human-human interactions could also be re-analysed only using the nativelikeness judgements provided by the native listeners as a measure of proficiency, to verify that the tendencies found using the PC1 are also present when using nativelikeness ratings directly. Second, we compared our combined proficiency measure with the nativelikeness ratings of a specific critical contrast, /æ, ɛ/. The question is how well these nativelikeness ratings reflect the participants' general accentedness in the L2 (rather than just the nativelikeness of these two vowels). For this reason, using only the nativelikeness ratings for the analysis of the data would not be ideal. Lastly, previous literature has found that the reliability of self-rated L2 proficiency tends to be modulated by, for example, speakers' level of anxiety (Hashemi, 2011; MacIntyre & Gardner, 1991) and their L1 (Diamond et al., 2019). Thus, the conclusion of this study may hold for the Dutch population but the generalization to speakers with other L1s would need to be supported with further research.

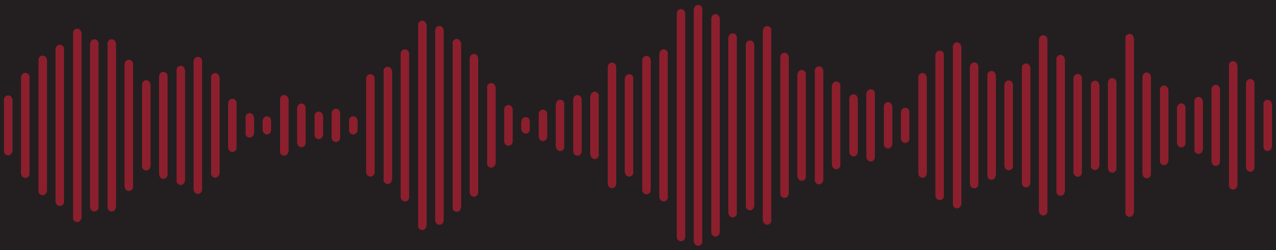
5.5.4. Conclusion

This study has confirmed that the non-native speakers' proficiency levels can affect the degree of adaptation they show in conversation. In fact, this study also shows that the effect of proficiency on non-native phonetic accommodation is different in human-human than in human-computer interactions. In the human-human setting, no differences were found between high and low proficiency speakers when they received exposure to the native-like pronunciation of the critical contrast and feedback on their own pronunciation. Interestingly, proficiency only seemed to modulate the degree of phonetic accommodation in the Control condition, where participants did not receive exposure or feedback, but where their attention was directed to the contrast by means of orthography. When non-native speakers interacted with a computer, the combination of the two factors, exposure and feedback, seemed to trigger more adaptation in the speakers with a lower proficiency level, which may be because accommodation could especially facilitate the communication with these learners. This study also has methodological implications: using a PCA encapsulating different proficiency measures proved to be the method that most reliably predicted the nativelikeness judgements produced by native speakers. In conclusion, in human-computer settings, low proficiency speakers were found to accommodate more to the interlocutor than high proficiency speakers. This finding supports the idea that accommodation, at least in human-computer settings, is an audience design phenomenon.



Chapter 6

General Discussion



The overarching goal of the current dissertation was to contribute to our understanding of non-native phonetic accommodation. Going a step beyond previous literature, the phenomenon of phonetic accommodation is here studied in spontaneous, yet controlled, interactive settings. Using the Ventriloquist paradigm (Chapter 2), this dissertation investigated how phonetic accommodation in non-native speakers is triggered by two conversation-intrinsic factors, namely the exposure a speaker receives to the interlocutor's pronunciation and the feedback a speaker receives on their own pronunciation. The effects of these two factors were studied from an acoustic (Chapter 3) and a perceptual perspective (Chapter 4). In addition, I investigated to what extent the effects of these two factors, exposure and feedback, were modulated by the type of interlocutor, human or a computer, (Chapter 3) and by the speaker's proficiency level in the L2 (Chapter 5).

This chapter discusses the general implications of the studies described in the previous chapters. Section 6.1 describes and discusses the methodological innovations presented in this dissertation. Section 6.2 discusses the theoretical questions addressed in this dissertation as regards non-native phonetic accommodation and L2 learning in interactions. Section 6.3 includes a general conclusion.

6.1. Methodological contributions

The current dissertation presents a number of methodological contributions to the field of phonetic accommodation and L2 learning. In particular, Chapters 2 and 4 describe two methodological innovations for the study and assessment of phonetic accommodation. Chapter 2 presents a new experimental paradigm developed for the study of phonetic accommodation in controlled, yet spontaneous settings, which is used in Chapter 3. Chapter 4 presents the use of Automatic Speech Annotation (ASA) technology to evaluate phonetic accommodation and L2 sound learning in terms of intelligibility. In addition, in Chapter 5, methodological conclusions can be drawn from the approach used to evaluate nativelikeness of L2 learners.

6.1.1. The Ventriloquist Paradigm

Chapter 2 presented and validated a new methodology, the Ventriloquist Paradigm, developed to study interactions in ecologically valid, yet controlled settings. First, the effectiveness of the methodology in making participants believe they were having a live conversation was assessed. The results showed that approximately 80% of the participants believed that they were interacting with a human confederate live and reported not noticing that the speech of the interlocutor was pre-recorded. Second, this chapter investigated whether this paradigm led to more spontaneous interactions than the use of computer-based settings. In order to do so, two versions of the paradigm were compared: one where participants interacted with a human confederate sitting in the same room making eye-to-eye contact, and a version of the paradigm where participants interacted with a computer. The participants who interacted with the human confederate talked significantly more and longer than those who talked to the computer. In other words, the Ventriloquist paradigm elicited more interaction from the participants than the setting with the computer interlocutor.

The findings of this chapter have implications for the methodology used to study interactions between human interlocutors. Previous literature studying phonetic accommodation in laboratory settings tends to sacrifice the spontaneity of the interaction in order to maintain experimental control over the input the participants receive (Babel et al., 2013; Cohn et al., 2019; Eijk et al., 2019; Lewandowski & Nygaard, 2018; Pardo et al., 2012, 2013; Rojczyk, 2013). In those studies, the interaction is typically simplified to low demanding tasks (Namy et al., 2002; Pardo, 2006) and, crucially, the interlocutor is drastically dehumanized, usually presented as a voice coming from a computer (e.g., Babel et al., 2013; Cohn et al., 2019; Eijk et al., 2019). The idea of using computer-based approaches to study human-human interactions lays on the Computers As Social Actors framework (Nass & Moon, 2000). According to this framework, humans tend to treat technology, such as computers, similarly to how they treat other humans. However, the generalizability of the results from studies using computer-based approaches to human-human interactions is questionable.

The findings from Chapter 2, together with the theoretical implications of chapters 3 and 4 (discussed in section 6.2) challenge the idea that computer-based settings can be used as a proxy to study interactions between humans. The patterns found in this study fall in line with previous research showing that in fact there are clear differences in how humans interact with other humans and with computers (Branigan et al., 2010; Burnham et al., 2010; Cohn et al., 2019; Raveh et al., 2019). Chapter 2 differs from previous literature investigating the differences between human-human and human-computer interactions in two respects. First, in this comparison the ecological validity of the conversations is preserved. The task on which the interaction was based moved away from simple, repetition-based tasks, such as shadowing tasks. Here a puzzle-solving task, based on the exchange of information between the interlocutors was used to elicit a conversation. Second, the human confederate was presented as a real human sitting in the same room as the participant, instead of as a virtual avatar (Staum Casasanto et al., 2010) or a video-presented confederate (Burnham et al., 2010). The findings in Chapter 2 show that the differences between the way humans interact with other humans and with computers also apply in spontaneous settings closer to everyday interactions. In fact, presenting the interlocutor face to face with the participants was found to be crucial. As described in Chapter 2, a version of the paradigm where the interlocutor was not present in the same room but in a different recording booth was not found to be as successful in making participants believe they were interacting with another human participant.

The differences found between the human-human and human-computer interactions indicate that the Ventriloquist Paradigm setting is a better suited methodology to study spontaneous interactions between human interlocutors than computer-based methodologies. The computer-based version used in this comparison resembles the Wizard of Oz paradigm (Riek, 2012), only differing in the implementation of a wide variety of spontaneous speech categories that help make the interaction spontaneous and more ecologically valid. This computer-based version can be a useful methodology for the study L2 sound acquisition and training in more interactive computer-based settings. In addition, given that the speech of virtual assistants is becoming more similar to human expressive vocal patterns (Cohn & Zellou, 2019; Skerry-Ryan et al., 2018; Tahon et al., 2020), this computer version of the paradigm can also help investigate how humans interact with recently developed virtual assistants.

Further research can explore new ways to improve the efficacy and ecological validity of the Ventriloquist paradigm. For example, since the first implementation of the Ventriloquist paradigm, described in Chapter 2, a new version of the paradigm has been developed. (Ye, 2020), a master thesis written under my supervision, introduced a newer version of the methodology used to study production and perception adaptations in German learners of English. This version, which successfully persuaded 90% of the participants that the confederate was speaking live, differed from the original paradigm in two different aspects. First, the recordings of the interlocutor's speech were designed to sound less artificial. The new recordings included repetitions, false starts and disfluencies in the speech to increase the spontaneity. Second, unlike in the first version of the paradigm, the order of the trials was fixed in Ye (2020). Thus, the speech used to describe the trials could include reference to previous trials. As regards the ecological validity, further research could explore the implementation of more interactive tasks, for example Diapix task (Van Engen et al., 2010) or the Map task (Anderson et al., 1984). Given that both tasks have been widely used in spontaneous interactions, designing the script for confederate's speech would be possible by thoroughly investigating previous corpora.

In this dissertation, the paradigm was used to study the learning of one difficult L2 vowel contrast, but further research could also broaden the scope of the implementation of this paradigm. For instance, Ye (2020) used the paradigm to study the learning of not only vowels but also consonant sounds during interaction. The scope could be further expanded to study the learning of prosodic, syntactic or lexical levels of the L2. Furthermore, this paradigm is not restricted to the study of speech production, but can also be used to study speech perception. Felker et al. (2021) used this paradigm to study which learning mechanisms were more effective in trigger L2 sound perception. Interestingly, Ye (2020) used this paradigm to study German learners of English L2 sound learning in terms of both production and perception.

6.1.2. Artificial intelligibility assessment

In Chapter 4, following previous literature (e.g., Cucchiarini et al., 2012; Escudero-Mancebo et al., 2015; Franco et al., 2010; I. Kim, 2006; Li et al., 2017; Van Nuffelen et al., 2009), Automatic Speech Recognition technology was used to automatize the intelligibility judgements on the speech of non-native speakers. In order to automatically produce the intelligibility judgements, I used an automatic annotator, Montreal Forced Aligner (MFA) (McAuliffe et al., 2017) with a modified version of the Carnegie Mellon University (CMU) dictionary (Weide, 1994). MFA performed a very similar task to that of the human listeners: based on the acoustic signal (for instance of the utterance “bad”), MFA chose between the English native-like (/bæd/) or the Dutch-accented pronunciation (/bɛd/) in the annotation of the utterance. Importantly, unlike previous literature validating the use of automatic annotators (e.g., van Nuffelen et al., 2009), the performance of MFA was compared not only to a small set of trained experts but also to a large pool of listeners, including over 100 native speakers of English.

The comparison between the automatically generated intelligibility judgements and the ones collected from native listeners indicate that artificial intelligibility judgements are similar but not as sensitive as human judgements. Similarly to previous literature (e.g., Kim et al., 2006; Cucchiarini et al., 2010), the artificial judgements were found to be correlated to the human judgements. In fact, in our study the overall intelligibility rating that a learner received from the human listeners was found to be strongly correlated to the artificial judgement generated using MFA (r = .73). However, the effects of exposure and feedback on phonetic accommodation appeared to differ depending on whether they were based on the artificial or the native listeners' judgements. This finding indicates that ASA systems could have a different degree of sensitivity than human listeners in their judgements of intelligibility.

In conclusion, the use of artificial intelligibility judgements comes with some advantages. These judgements can reflect judgements collected from human listeners and they are immensely more efficient than collecting human ratings. Collecting perceptual judgements on speech is difficult and time-consuming task. For example, the presentation of stimuli to listeners needs to be designed in a thoughtful way, taking into account, for example, human limitations on attention and unpredictable human behavior. The use of artificial judgements, such as the ones generated by MFA in Chapter 4, requires a much simpler design –no lists organizing the stimuli are required– and virtually no restrictions apply to the selection of the stimuli. There are virtually no restrictions on the amount and characteristics of the utterances that can be processed by MFA. For example, the words selected for the perception experiment with the human listeners needed to part of an /æ, ɛ/ word-pair that existed in English, were frequent and monosyllabic. The methodology used to generate the artificial judgements, by including modifications in the dictionary entries, does not impose any of those restrictions that apply to the stimuli for human listeners.

However, some limitations need to be taken into account too. The ASA system was not found to be as sensitive as the human ratings in perceiving differences in the pronunciation. Thus, although the general intelligibility feedback was found to be strongly correlated, acoustic cues may have different weights in the production of the intelligibility judgements by humans and by MFA. In addition, using automatically generated ratings instead of a large pool of raters in the study reported in Chapter 4 led to restriction of the number of observations and, therefore, the statistical power for the analysis. These findings are of course restricted to the sounds under study. No other sounds were included in the comparison.

6.1.3. Propficiency assessment reflecting nativelikeness

Chapter 5 presents a study where the role of L2 proficiency in phonetic accommodation was investigated. The definition and assessment of L2 proficiency can pose a challenge to researchers. Different methodologies have been used to assess L2 proficiency, such as the LexTale test (Lemhöfer & Broersma, 2012) or self-rated measures. These methodologies usually correlate well with overall proficiency measures, such as the Common European Framework for Reference (Council of Europe, 2001). The hypotheses included in Chapter 5 required an assessment of proficiency that not only reflected the general proficiency in the L2, with an indication of the cognitive resources used, but also the nativelikeness of the pronunciation of L2 sounds.

In Chapter 5, seven different proficiency measures were used, including an objective measure –LexTale score (Lemhöfer & Broersma, 2012)– and six self-rated proficiency measures including writing, listening, speaking and reading proficiency, strength of the non-native accent and how often the learner is identified as a non-native speaker. The different assessments were combined into a single proficiency measure using Principal Component Analysis. The components were compared to the nativelikeness ratings provided by human listeners on the same speech recordings collected in Chapter 4 and PC1, the component that was representative of all the measures used, showed the highest correlation to the nativelikeness judgements provided by the human listeners. In fact, a post-hoc analysis was run in order to determine whether any of the individual measures better correlated with the nativelikeness ratings. This analysis showed that PC1 was the measure that highest correlated with the nativelikeness ratings (see table in appendix 34), even more than the self-rated judgements directly related to the nativelikeness, namely strength of the non-native accent and how often the learner was identified as a non-native speaker.

This finding has methodological implications for the assessment of proficiency, in particular of proficiency measures that reflect the nativelikeness in the L2 pronunciation. First, the fact that PC1 showed a higher correlation than the other components, which were mostly representative of smaller subsets of the measures, indicates that the combinatory measure can more reliably predict nativelikeness than individual measures in this population of L2 learners. Second, this finding also shows that the nativelikeness of the L2 pronunciation can be somewhat reliably predicted using self-rated proficiency measures that are easier to collect. This second aspect is indeed interesting for the methodology used to evaluate L2 proficiency because the collection of these self-rated proficiency measures is evidently more efficient than collecting nativelikeness judgements on learners' pronunciation. However, these findings come with some limitations. Previous literature has found that the reliability of self-rated measures is linked to the L1 of the speakers (Diamond et al., 2019). The tendencies found in this study with Dutch learners of English could differ with learners of a different language or a different L1.

6.2. Theorical contributions

6.2.1. Non-native phonetic accommodation

In Chapter 3, 4 and 5, I investigated which mechanisms underlie non-native phonetic accommodation. In order to do so, I investigated interactions between Dutch learners of English and a native speaker of English and analyzed the learners' pronunciation of the /æ, ɛ/ contrast before and after the interaction. Dutch learners of English, even advanced learners, tend to merge the two vowels into an in-between sound (Elsendoorn, 1985; Wang & van Heuven, 2006), so accommodation was interpreted as any adaptation applied to L2 learners' pronunciation with the purpose of unmerging these two vowels. I studied whether accommodation in non-native speakers was triggered by two factors: exposure to the native speaker's pronunciation of a given contrast, the English /æ, ɛ/ pair, and feedback on the same contrast indicating that the interlocutor was struggling to understand the message. The following sections describe the main findings as regards the effect of exposure and feedback in human-human settings (Chapters 3 and 4), how these effects are different in human-computer settings (Chapter 3) and the role the learner's proficiency plays in how effective those factors are in triggering phonetic accommodation in both human-human and human-computer interactions (Chapter 5).

6.2.1.1. The effect of exposure on non-native phonetic accommodation

Chapters 3 and 4 of this dissertation investigated whether receiving exposure to the native pronunciation during an interaction triggered phonetic accommodation. The exposure consisted in including twelve minimal word pairs words with the critical contrast /æ, ɛ/ in the speech of the interlocutor. In previous literature, the effect of exposure to the interlocutor speech is rarely disentangled from the effect of just participating in an interaction. In Chapters 3 and 4, the effect of exposure to the interlocutor's pronunciation was isolated from the effect of participating in the interaction by comparing the phonetic accommodation in participants who received exposure to the interlocutor's pronunciation of the critical contrast, to a control group of participants who participated in an identical interaction but did not receive exposure to the critical contrast during the interaction. Therefore, the effect of exposure on phonetic accommodation can be successfully assessed separately from the experience of just participating in the interaction.

Chapters 3 and 4 assess phonetic accommodation from an acoustic and a perceptual perspective, respectively, following previous literature (Pardo et al., 2013). In Chapter 3, accommodation, as measured with the height (first formant) and frontness (second formant) of the /æ, ɛ/ sounds, was found to be boosted by the presence of exposure to the native pronunciation of the contrast. These adaptations were not found in the group of participants used as a control, who did not receive exposure at all during the interaction. In Chapter 4, the presence of exposure led to an increase in the intelligibility of the pronunciation of /æ/, which was not found in the participants assigned to the control condition. Participants who received exposure improved the nativelikeness of their pronunciation of /æ/similarly to participants in the control condition, who did not receive any exposure to the contrast during the interaction.

The findings of these chapters contribute to the study of phonetic accommodation (Costa et al., 2008; Pickering & Garrod, 2004) in three different ways. First, the results from these two chapters help understand the mechanisms underlying accommodation. From a theoretical point of view, investigating the effect of exposure on phonetic accommodation is crucial in order to determine whether this phenomenon is an automatic mechanism resulting from priming, as proposed in Pickering and Garrod (2004) and Costa et al. (2008). From this perspective, phonetic accommodation is understood as a speaker-driven phenomenon: by exposing a speaker to a certain cue, the higher level of activation of the cue facilitates the future retrieval and use of the cue (Pickering & Garrod, 2004). The findings from chapters 3 and 4 fall in line with previous literature showing that speakers adapt their speech in interaction as the result of receiving exposure (Babel, 2010, 2012; Babel et al., 2013; Berry & Ernestus, 2018; Eijk et al., 2019; Gijssels et al., 2016; Levitan & Hirschberg, 2011; Pardo et al., 2012; Staum Casasanto et al., 2010). Therefore, these findings provide evidence for the account of accommodation as an automatic mechanism.

Second, this dissertation helps expand the scope of the research on phonetic accommodation as an automatic mechanism to also include non-native contexts. Following the priming account, non-native speakers have been theorized to be less likely to show accommodation as the result of priming (Costa et al., 2008). Using an L2 is a cognitively demanding task (Lecumberri et al., 2010); therefore, non-native speakers were assumed to have fewer cognitive resources available to focus on the input from the interlocutor and be primed. The findings from these chapters, however, show that Dutch learners of English can be primed by exposure to a sound contrast in the speech of the interlocutor during a spontaneous conversation. This effect was found even though the participants were engaged in a demanding task, i.e., using an L2 in the context of an interaction.

Third, unlike previous literature studying the effect of exposure on phonetic accommodation (Babel, 2010; Babel et al., 2013; Babel & Bulatov, 2012; Berry & Ernestus, 2018; Staum Casasanto et al., 2010; Eijk et al., 2019; Gijssels et al., 2016; Pardo et al., 2012), the current dissertation studied the effect of this factor combining experimental control and spontaneity in a human-human interaction. The experimental control allowed to disentangle the effect of exposure to the specific sounds from the overall effect of participating in the interaction. The ecological validity of the setting, with a communicative task and a human interlocutor facilitates the generalizability of the results to everyday interactions between humans (see section 6.1.1).

6.2.1.2. The effect of feedback on non-native phonetic accommodation

Chapters 3 and 4 investigated the effect of receiving feedback implying misunderstanding on phonetic accommodation. The feedback consisted in the confederate systematically pretending to misunderstand the participants' pronunciation of words including the /æ/ vowel (e.g., "bad") for minimal pair words including the /ɛ/ sound (e.g., "bed") in twelve instances during the interaction. Importantly, the use of pre-recorded speech for the confederate's input enabled assessing the effect of feedback on phonetic accommodation separately from exposure to the critical vowel contrast.

Chapter 3 showed that participants adapted their pronunciation of the critical vowels, in terms of the height and frontness, after receiving feedback from the interlocutor, even when they received no exposure to the native pronunciation of the sounds. The results from Chapter 4 showed that the presence of feedback led to an increase in the intelligibility of learners' pronunciation, compared to a control condition where participants did not receive feedback or exposure. The nativelikeness, in contrast, was not boosted by the presence of mere feedback: participants who received only feedback increased their nativelikeness after the interaction similarly to the participants in the control condition, who did not receive feedback or exposure.

The findings from chapters 3 and 4 combined have several implications on our understanding of phonetic accommodation. Investigating to what extent feedback can trigger phonetic accommodation is essential in order to understand whether accommodation is a listener-driven phenomenon (Bell, 1984). Parallel to the account of accommodation as an automatic mechanism, this phenomenon has also been argued to be driven by the speaker's eagerness to ensure communicative success (Bell, 1984). According to this view, speakers accommodate their speech to the interlocutor not because they are primed by the input they receive but because they want to ensure the interlocutor's understanding (Bell, 1984). The feedback included in chapters 3 and 4 is a relevant factor within this approach because it signals to the speaker that the interlocutor is struggling to understand the message. These chapters follow previous literature (e.g., Burnham et al., 2010; Gessinger et al., 2019) in providing evidence that the presence of feedback implying misunderstanding from the interlocutor can trigger phonetic accommodation.

Critically, in these studies feedback was not studied separately from exposure: participants received exposure to the interlocutor's pronunciation of the contrast that was subject to feedback (e.g., Burnham et al., 2010; Gessinger et al., 2019). Nevertheless, from L2 classroom studies, it is clear that the effect of feedback differs drastically when presented in isolation and when presented in combination with "positive evidence", i.e., exposure to the cue targeted by the feedback (e.g., Adams et al., 2011; Dlaska & Krekeler, 2013; Mackey & Gass, 2006; ; Saito & Lyster, 2012). Not isolating feedback from exposure to the interlocutor's pronunciation hinders the conclusions drawn from these studies on phonetic accommodation: it is not clear to what extent the accommodation is the result of receiving feedback or the result of receiving exposure. Chapter 3 and 4, in contrast, thus, show that speakers can adapt their pronunciation as the result of receiving feedback, regardless of the presence of exposure to the interlocutor's speech.

Also unlike previous literature (Burnham et al., 2010; Gessinger et al., 2019), the effect of feedback in accommodation was assessed in a spontaneous setting involving two human interlocutors present in the same room, where the participant (the non-native speaker) believed they were having a spontaneous conversation. The use of a spontaneous setting with a human interlocutor enables the generalizability of the results to every-day interactions (see section 6.1.1.).

6.2.1.3. The effect of exposure and feedback on non-native phonetic accommodation

As discussed in the previous section, the effects that exposure and feedback may have on the adaptations speakers may apply to their speech in the course of an interaction are rarely disentangled (e.g., Burnham et al., 2010; Gessinger et al., 2019). Chapters 3 and 4, using pre-recorded speech, investigated the effect of feedback presented together with exposure on phonetic accommodation in Dutch learners of English interacting with a human confederate.

In Chapter 3, although the two factors in isolation had triggered adaptations in the learners' pronunciation (as described in the previous sections), the combination of the two factors was not found to trigger any accommodation at all in the height and frontness of the vowels under study. In Chapter 4, the results showed that the combination of the two factors together improved the intelligibility but not the nativelikeness of the learners' pronunciation in the L2. These findings, related to the intelligibility of the participants, follow previous literature showing that phonetic accommodation can be triggered by exposure (Eijk et al., 2019; Gijssels et al., 2016; Levitan & Hirschberg, 2011; Staum Casasanto et al., 2010) and feedback (Burnham et al., 2010; Gessinger et al., 2019). In this case, the adaptations participants applied to their pronunciation did not aim to increase the similarities between them and the interlocutor, i.e., the nativelikeness, but were aimed to increase the intelligibility of their speech.

These findings, thus, highlight the importance of using a combination of acoustic measures and perceptual judgements when measuring phonetic accommodation. As discussed in Pardo et al. (2013), the use of acoustic analysis can provide fine-grained measurements of accommodation but often offer a narrow scope for analysis. In fact, the perceptual measures used in Chapter 4 seem to be a more comprehensive assessment of accommodation, which not only provides a broader scope of different acoustic dimensions used by the listeners to produce their judgements but also confirms that the adaptation can indeed be perceived by the interlocutor. Thus, the results of these two chapters combined supports the use of both acoustic and perceptual measurements to assess phonetic accommodation.

As regards the mechanisms underlying phonetic accommodation, the findings of the combination of the two factors provide mixed evidence for phonetic accommodation to be both an automatic process (Pickering & Garrod, 2004) –resulting from exposure–, and an audience-design process (Bell, 1984) –resulting from feedback. The acoustic analysis used in Chapter 3 provides no evidence for the combination of the two factors to trigger phonetic accommodation. A possible explanation is that the combination of the two factors could have been interpreted to be confusing for participants of the experiment, given that the feedback was kept consistently corrective. If participants were confused, they could have noticed the consistency of the corrective feedback and, therefore, stopped attempting to accommodate their pronunciation. However, the results described in Chapter 4 show that the two factors together not only triggered adjustments in L2 speakers' intelligibility but in fact the increase was significantly greater than when the factors, exposure and feedback, were presented in isolation. This finding indicates that the feedback together with exposure triggered different accommodating behavior than feedback in isolation.

Taking into account the results from the intelligibility and the nativelikeness judgements together, the patterns found in Chapter 4 when participants received both exposure and feedback seem to fit better with the audience-design explanation of accommodation (Bell, 1984) than with a priming account (Pickering & Garrod, 2004). If accommodation is the result of priming, the adaptations could be expected to result in participants sounding more similar to the interlocutor (i.e., more native-like) after the interaction. In the other conditions (control condition, OnlyExposure and OnlyFeedback), participants overall improved their nativelikeness. However, when participants received exposure and feedback together, the overall improvement in the nativelikeness disappeared. In other words, participants did not sound more similar to the interlocutor in the post-test than in the pre-test. The adaptations participants in this condition made to their pronunciation seemed to be aimed at ensuring understanding, thus the increase in intelligibility, but did not lead to an increase in nativelikeness. Thus, in this condition the result seems to support an audience-design perspective to explain phonetic accommodation (Bell, 1984) rather than a priming perspective (Pickering & Garrod, 2004).

6.2.2. Phonetic accommodation in computer settings

In Chapter 3, I investigated whether phonetic accommodation is dependent on the type of interlocutor. Here, a direct comparison is drawn between phonetic accommodation in human-human settings (described in the previous section) and in human-computer settings, based on identical settings only differing in the type of interlocutor (a human confederate or a computer). In particular, the study presented in Chapter 3 compared the effects of exposure and/or feedback on phonetic accommodation in human-computer and human-human interactions.

The results of this study show that the presence of exposure to the native-like pronunciation of the critical contrast, /æ, ɛ/ during the interaction triggered more phonetic accommodation, measured as changes in the height and frontness of the critical vowels, in the setting where speakers interacted with a human confederate than in the setting where they interacted with a computer, following previous literature on the effect of feedback (Cohn et al., 2019). The presence of feedback in isolation triggered more accommodation in the human-human setting than in the human-computer setting, also following previous literature on the effect of feedback (Burnham et al., 2010). The combinations of the two factors, exposure and feedback, in contrast, triggered more adaptations in the human-computer setting than in the human-human setting, also in line with previous literature on the effect of feedback (Burnham et al., 2010). Although the combination of exposure and feedback could have been confusing in the human-human setting; in the computer setting, on the other hand, participants may have interpreted the consistency of the corrective feedback to be part of the limitations of technology.

The results from the study described in Chapter 3 shows that the factors underlying phonetic accommodation, such as exposure and feedback, can be processed and used in drastically different ways when the interlocutor is presented as a human or as a computer. The differences found in the patterns of results help us understand what drives phonetic accommodation in human and computer interactions and partly validate previous literature discussing the differences between human-human and human-computer accommodation (Branigan et al., 2010).

Branigan et al. (2010) explains that accommodation in human-human and human-computer interactions are motivated by different underlying goals. Whereas human-human accommodation is a phenomenon driven by unmediated mechanisms such as priming, human-computer accommodation is a more strategic process driven by the communicative success of the conversation. Following this distinction, speakers are expected to be more likely primed by other human interlocutors than by computers. Similarly to Cohn et al. (2019), the results of this dissertation also show that exposure to the interlocutor's speech triggers more accommodation when it comes from a human than when it comes from a computer. Thus, this finding supports the claim that accommodation as the result of priming is more likely to take place when the interlocutor is a human than when the interlocutor is a computer (Branigan et al., 2010).

In interactions with computers, accommodation has been argued to be a more strategic process aimed to ensure the communicative success of the interaction (Branigan et al., 2010). In this dissertation, this claim was investigated with the implementation of feedback during the interaction. The feedback would let participants know that the communicative success was threatened. The current dissertation found evidence supporting this claim, but only in the condition where participants received the feedback implying misunderstanding from the interlocutor together with exposure to the interlocutor's pronunciation. The participants in the human-computer setting accommodated more to the interlocutor after receiving exposure and feedback than the participants who interacted with a human confederate. This finding falls in line with previous research (e.g., Burnham et al., 2010; Gessinger et al., 2019), where the effect of feedback on phonetic accommodation was stronger in human-computer interactions than in human-human interactions.

Following Branigan et al. (2010), feedback in isolation should be expected to trigger more accommodation in human-computer settings than in human-human settings. Interestingly, the effect of feedback in isolation was not found to follow the same pattern as the effect of feedback together with exposure. The participants who received feedback (but no exposure) from the human interlocutor showed more accommodation than those who interacted with the computer. Thus, unlike in the condition where feedback was combined with exposure, no evidence was found in this condition for the audience-design account of accommodation to be more relevant in human-computer settings than in human-human settings (Branigan et al., 2010).

The findings from this dissertation show clear differences between accommodation to humans and to computers. The differences found in terms of the effect of exposure and feedback highlight that the role of the interlocutor is crucial in order to understand the weight different factors have on phonetic accommodation. Consequently, the findings of Chapter 3 argue against the use of computer-based approaches to study human-human interaction (as also discussed in 6.1.1). In addition, these findings also indicate that the definition of phonetic accommodation, as an automatic process or an audience design phenomenon, is strictly dependent on the type of interlocutor involved in the interaction.

6.2.3. The effect of proficiency on phonetic accommodation

This dissertation also aimed to explain how individual differences may modulate the degree of phonetic accommodation displayed by speakers. In Chapter 5, I investigated whether L2 proficiency mediated the effect of exposure and feedback on phonetic accommodation. Since phonetic accommodation has been found to differ depending on the type of interlocutor (Branigan et al., 2010; Chapter 3), this chapter compares the effect of proficiency on accommodation in human-human and human-computer settings.

The results of the study presented in Chapter 5 show differences in the role of proficiency in L2 accommodation depending on the type of interlocutor. In the human-human setting, proficiency was not found to modulate the effects of exposure or feedback on phonetic accommodation at all. However, in the control condition, where participants did not receive exposure or feedback, high proficiency speakers were found to show more accommodation than low proficiency speakers. In the human-computer setting, in contrast, proficiency was found to modulate the effect of exposure and feedback (in combination) on phonetic accommodation in the opposite direction: L2 learners with a lower proficiency level were found to accommodate more than high proficiency learners in this condition.

Following the account of accommodation as an automatic mechanism resulting from priming (Pickering & Garrod, 2004), high proficiency speakers could be expected to accommodate more than low proficiency speakers (Berry & Ernestus, 2018; Costa et al., 2008). Costa et al. (2008) argues that speakers with fewer cognitive resources available will be less likely to show accommodation than those with more resources available. Consequently, non-native speakers are expected to accommodate less than native speakers (Costa et al., 2008). Following this perspective, high proficiency L2 learners, who can process L2 speech more fluidly (Lecumberri et al., 2010), could be expected to accommodate more than low proficiency learners (Berry & Ernestus, 2018). Unlike previous literature (e.g., Berry & Ernestus, 2018), no evidence was found in Chapter 5 supporting this view: the effect of exposure in isolation was not found to be modulated by the speaker's proficiency level in either of the settings.

From an audience-design perspective on accommodation (Bell, 1984), speakers could be expected to accommodate more when the communicative success is more at risk. Low proficiency speakers tend to have a more accented pronunciation in the L2 than high proficiency speakers. Following Bell (1984), low proficiency speakers could be expected to accommodate more to a native speaker interlocutor because their pronunciation could pose a bigger threat to the communicative success. The findings described in Chapter 5 partially provide evidence for this perspective in the human-computer setting. The presence of feedback during the interaction was found to trigger more accommodation in low proficiency speakers than in high proficiency speakers. This finding could be interpreted as evidence for the willingness to ensure communicative success to be a factor facilitating phonetic accommodation. However, this tendency was only found in Chapter 5 when exposure was presented in combination with the feedback and not when feedback was presented in isolation. Thus, it was not only the threat posed to the interaction that made participants changed their pronunciation. The accommodating behavior was also triggered partially by the presence of exposure.

In general, the findings of this chapter indicate that the proficiency level of the speakers in the L2 can play a role in how much they can adapt their speech during an interaction. No evidence was found to support the explanation of accommodation based on the cognitive resources available (Berry & Ernestus, 2018; Costa et al., 2008) and partial evidence was found to support the audience-design perspective (Bell, 1984), only in the human-computer setting.

6.2.4. How does non-native accommodation work in non-native contexts?

In general, the findings of the chapters presented in this dissertation indicate that phonetic accommodation can indeed take place in non-native contexts. I showed that phonetic accommodation in non-native speakers can be both an automatic process and an audience design phenomenon: accommodation can take place both as the result of receiving exposure to the specific sounds of the L2 and also as the result of receiving feedback implying misunderstanding from the interlocutor. Thus, this dissertation argues for a definition of accommodation combining both perspectives.

In addition, this dissertation clearly shows that phonetic accommodation in human-human settings is drastically different from human-computer settings. The results of this dissertation provide evidence for the presence of accommodation in human-computer setting, following the approach that humans treat computers like they would treat other humans (Nass & Moon, 1998). However, the clear differences between the settings shows that computer-based approaches cannot be generalizable to everyday interactions between human speakers.

The results from the study investigating the role of the L2 proficiency provide evidence for phonetic accommodation to be dependent on cognitive resources in some settings but dependent on ensuring the communicative success in others. This dissertation suggests that the different definitions used to explained accommodation drastically depend on the context where the conversation is taking place.

This dissertation focused on investigating phonetic accommodation from two different perspectives: the audience design account and the automatic mechanism approach. Further research could also take into account the social dimension of this phenomenon. For example, future studies could use other accents for the voice of the confederate other than a dialect of English that is to be standard. It would be interesting to investigate whether the effects of exposure and feedback found in this dissertation could also be found when the voice includes a variety of English with different social connotations. Furthermore, future research could also explore using the Ventriloquist paradigm to study phonetic accommodation between native speakers of the same language, instead of non-native speakers. This project would help understand to what extent the effect of feedback and exposure on phonetic accommodation may differ in L1 and L2 speakers in phonetically controlled, yet spontaneous settings.

6.2.5. L2 learning

The current dissertation also has shed new light on the process of L2 sound acquisition. Dialogues are a common setting for L2 learning to take place (Ellis, 1999), but so far little is known about how this process may develop in interactions. Phonetic accommodation could be the mechanism underlying L2 sound acquisition. The adaptations that non-native speakers apply to their speech when interacting in the L2 could make become permanent changes in their L2 sound systems and lead to L2 learning. L2 acquisition has been vastly studied and documented in classroom settings (e.g., Loewen & Philp, 2006; Lyster, 1998; Mackey et al., 2007; Mackey & Gass, 2006, among others) and computer-based settings (e.g., Cucchiarini et al., 2012; Nushi et al., 2017; Saito & Lyster, 2012). Previous studies have identified different mechanisms that can be used in order to foster the L2 learning in these settings (for exposure to the native pronunciation see Adams et al., 2011; Dlaska & Krekeler, 2013; Mackey & Gass, 2006; Saito & Lyster, 2012; for feedback see Adams et al., 2011; Cucchiarini et al., 2012; Dlaska & Krekeler, 2013; Lee & Lyster, 2017; Nassaji, 2009; Saito & Lyster, 2012). However, little is known about what fosters this process in interactions.

The current dissertation aimed to study phonetic adaptations during the interaction that could be maintained in the L2 sound systems of the learners. For this reason, in Chapters 3, 4 and 5, the focus of study is on the production before and after the interaction, instead of the production by the learners during the interaction. I aimed to capture the changes in the pronunciation that endure in a short-term timeframe. The results of the studies presented in this dissertation can therefore be interesting from an L2 learning perspective too.

In particular, this dissertation explains how different conversation-intrinsic factors, such as exposure to the native-like pronunciation and feedback on the learners' pronunciation, can affect

L2 sound learning. First, I show that L2 learning in interactions can be boosted by exposure to the native-like pronunciation. Second, the presence of feedback during the interaction was also found to trigger L2 sound learning. However, the effect of feedback was dependent on whether it was presented with or without exposure. Without exposure, feedback was found to trigger learning of the L2 sounds involved in the feedback. With exposure, however, the effect of feedback was found to trigger more intelligible, but not more native-like, pronunciation of the contrast under study.

This study also helps understand how similar L2 learning is in computer-based settings and human-human interactions and how exposure and feedback may work in these settings. Since L2 learning can take place in different contexts, such as classroom settings, computer-based settings or human-human interactions, knowing what triggers L2 learning in these contexts is essential in order to understand how to optimize learning in these different settings. The findings from the human-human and human-computer comparison presented in Chapter 3 shed some light on how learning may differ in these settings. L2 sound learning may benefit more from human-based exposure than computer-based exposure. Thus, when developing methodologies to teach L2 sounds in interaction, implementing human-like interlocutors will be more beneficial than computer-based learning.

The effect of feedback without exposure was found to be similarly effective in the two settings at triggering L2 learning. However, the implementation of feedback with exposure drastically differed in the two settings. The combination of exposure and feedback was less effective in the human-human settings perhaps because it was interpreted as confusing. In the human-computer setting, however, it was found to be effective in triggering changes in the production of the vowels. This finding thus shows that learners tend to be more understanding and accepting of the feedback when it comes from a computer than when it comes from a human.

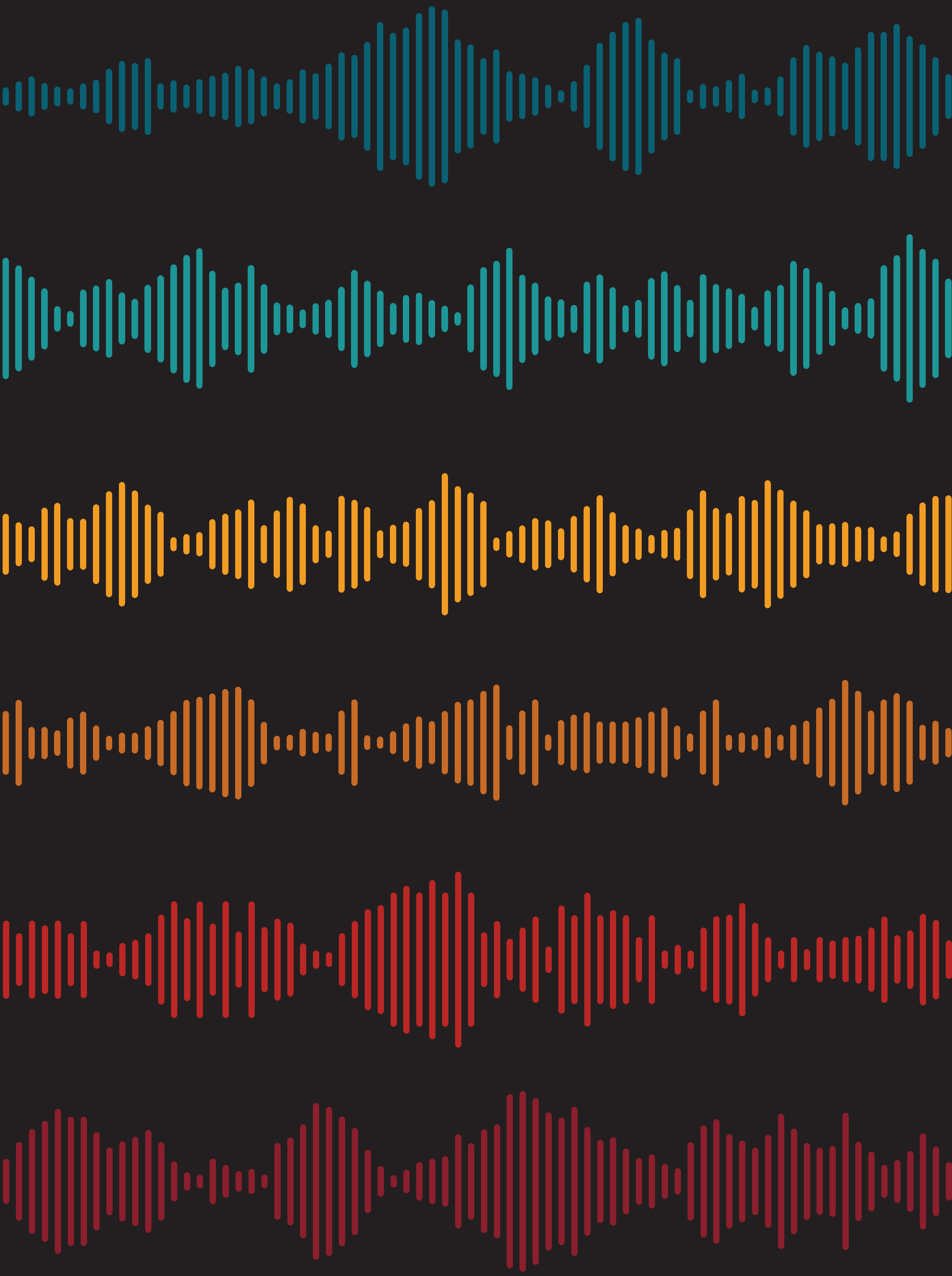
Chapter 5 shows that the capacity for L2 sound learning in interaction can be modulated by the degree of proficiency in the L2.

In particular, in the human-computer setting, participants with a low proficiency in the L2 benefited more from receiving exposure to the native-like pronunciation and feedback on their pronunciation than high proficiency learners. This finding indicates that when designing L2 learning methodology, taking into account the level of proficiency of the learners is essential for the optimalization of the methodology.

6.3. Conclusion

This dissertation has contributed to our resources to study phonetic accommodation and our understanding of this process in non-native speakers. As for the resources used to study phonetic accommodation, this dissertation has yielded new research methodology which will benefit researchers investigating not only phonetic accommodation but also L2 learning. In particular, a new experimental paradigm has been developed which successfully combines natural settings and experimental control over the input participants receive during the interaction. This methodology provides researchers with more ecologically valid situations to study interactions. Most importantly, the results of the studies included in this dissertation, where this methodology was used, highlight the importance of studying communicative processes, such as phonetic accommodation, in ecologically valid situations that enable a more direct generalization to every-day interactions. In addition, this dissertation provides an example of how Automatic Speech Annotation technology can be used to rate L2 learners' pronunciation in an efficient way.

From a theoretical point of view, I successfully documented the presence of this phenomenon in non-native speakers using a spontaneous, yet controlled, methodology. Accommodation was both assessed from an acoustic point of view and a perceptual perspective; thus, the phonetic adaptations were perceivable by the human ear. Furthermore, I identified clear differences between accommodation in human-human and human-computer settings. These differences indicate that the definition of phonetic accommodation heavily depends on the type of interlocutor. The general tendency found suggests that accommodation as the result ensuring the communicative success is more relevant in the human-computer setting than in the human-human setting.



References

Adams, R., Nuevo, A. M., & Egi, T. (2011). Explicit and Implicit Feedback, Modified Output, and SLA: Does Explicit and Implicit Feedback Promote Learning and Learner-Learner Interactions? *The Modern Language Journal*, *95*, 42–63. https://doi.org/10.1111/j.1540-4781.2011.01242.x

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107. https://doi.org/10.1121/1.1795335

Aliaga-García, C. and Mora, J. C. (2008). Assessing the effects of phonetic training on L2 sound perception and production. In Rauber, A. S., Watkins, M. A., and Baptista, B. O. (Eds.). *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (pp. 10-27). Florianópolis: Federal University of Santa Catarina

Anderson, A., Brown, G., Shillcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge University Press.

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(1), 190–201. https://doi.org/10.1017/S1366728911000125

Auer, P., Barden, B., & Grosskopf, B. (1998). Subjective and Objective Parameters Determining 'Salience' in Long-term Dialect Accommodation. *Journal of Sociolinguistics*, *2*(2), 163–187. https://doi.org/10.1111/1467-9481.00039

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, *55*(2), 231–248. https://doi.org/10.1177/0023830911417695

Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology*, *4*. https://doi.

org/10.3389/fpsyg.2013.00653

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, A. (1984). Language Style as Audience Design. *Language in Society*, *13*(2), 145–204.

Berry, G. M., & Ernestus, M. (2018). Phonetic alignment in English as a lingua franca: Coming together while splitting apart. *Second Language Research*, *34*(3), 343–370. https://doi.org/10.1177/0267658317737348

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer [Computer program]. Version 6.1.07, retrieved 26 November 2019 from http://www.praat.org/*.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303–328. https://doi.org/10.1017/S0142716400008912

Bohn, O.-S., & Flege, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Studies in Second Language Acquisition*, *14*(2), 131–158. https://doi.org/10.1017/S0272263100010792

Bradac, J. J., Mulac, A., & House, A. (1988). Lexical diversity and magnitude of convergent versus divergent style shifting-: Perceptual and evaluative consequences. *Language & Communication*, *8*(3–4), 213–228. https://doi.org/10.1016/0271-5309(88)90019-5

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, *42*(9), 2355–2368. https://doi.org/10.1016/j.pragma.2009.12.012

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Nass, C. I. (2003). Syntactic alignment between computers and people: The role of belief about mental states. *Proceedings of the 25th Annual Conference of the Cognitive*

Science Society, July 31 - August 2 2003, Boston, Massachusetts, 186–191.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, *64*(1), 74–95. https://doi.org/10.1080/17470218.2010.499174

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer- and human-directed speech before and after correction. *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*, 13–17.

Cenoz, J., & García-Lecumberri, L. (1999). The effect of training on the discrimination of English vowels. *IRAL - International Review of Applied Linguistics in Language Teaching*, *37*(4). https://doi.org/10.1515/iral.1999.37.4.261

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Christensen, R.H.B. (2019). *ordinal—Regression Models for Ordinal Data .* R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Cohn, M., Segedin, B. F., & Zellou, G. (2019). Imitating Siri: Socially-mediated vocal alignment to device and human voices. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1813-1817.

Cohn, M., Zellou, G. (2019) Expressiveness Influences Human Vocal Alignment Toward voice-AI. *Proceedings of Interspeech 2019*, 41-45, doi: 10.21437/Interspeech.2019-1368

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*(4), 528–556. https://doi.org/10.1080/01690960801920545

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge : Cambridge University Press, 2001.

Cucchiarini, C., Nejjari, W., & Strik, H. (2012). My Pronunciation Coach: Improving English pronunciation with an automatic coach that listens.

Language Learning in Higher Education, *1*(2). https://doi.org/10.1515/cercles-2011-0024

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. https://doi.org/10.1121/1.1810292

Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention Control and Inhibition Influence Phonological Development in a Second Language. *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics, 5.*

Delvaux, V., & Soquet, A. (2007). The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica*, *64*(2–3), 145–173. https://doi.org/10.1159/000107914

Diamond, L., Toro Bejarano, M., Chung, S., Ferguson, W., Gonzalez, J., Genoff Garzon, M., Mujawar, I., & Gany, F. (2019). Factors Associated With Accuracy of Self-Assessment Compared With Tested Non–English Language Proficiency Among Primary Care Providers. *Medical Care*, *57*(5), 385–390. https://doi.org/10.1097/MLR.0000000000001105

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*(1), 25–37. https://doi.org/10.1016/j.system.2013.01.005

Eijk, L., Ernestus, M., & Schrievers H. (2019). Alignment of pitch and articulation rate. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia 2019, pages 2690-2694. doi:10.1016/j.wocn.2019.02.004.

Elsendoorn, B. A. G. (1985). Production and Perception of Dutch Foreign Vowel Duration in English Monosyllabic Words. *Language and Speech*, 28(3), 231–254. https://doi.org/10.1177/002383098502800302

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. https://doi.org/10.1016/j.lingua.2012.12.006

Escudero-Mancebo, D., Camara-Arenas, E., Tejedor-Garcıa, C., Gonzalez-Ferreras, C., & Cardenoso-Payo, V. (2015). Implementation and test of a serious game based on minimal pairs for pronunciation training. *SLaTE* 125-130.

Eurobarometer, Special. (2012). *Europeans and their Languages*. European Comission.

Felker, E. (2021). *Learning second language speech perception in natural settings*. LOT. https://doi.org/10.48273/LOT0590

Felker, E., Troncoso-Ruiz, A., Ernestus, M., & Broersma, M. (2018). The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America*, *144*(4), EL304–EL309. https://doi.org/10.1121/1.5063809

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, 93(3), 1589-1608.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V., & Precoda, K. (2010). EduSpeak: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing*, *27*(3), 401–418. https://doi.org/10.1177/0265532210364408

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, *5*(1), 81–99. https://doi.org/10.1016/0885-2308(91)90019-M

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://CRAN.R-project.org/package=irr

Gessinger, I., Möbius, B., Andreeva, B., Raveh, E., & Steiner, I. (2019). Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments. *Interspeech 2019*, 301–305. https://doi.org/10.21437/Interspeech.2019-2445

Gessinger, I., Raveh, E., Le Maguer, S., Möbius, B., & Steiner, I. (2017, August). Shadowing Synthesized Speech-Segmental Analysis of Phonetic Convergence. In *Interspeech* (pp. 3797-3801).

References

Gijssels, T., Casasanto, L. S., Jasmin, K., Hagoort, P., & Casasanto, D. (2016). Speech Accommodation Without Priming: The Case of Pitch. *Discourse Processes*, *53*(4), 233–251. https://doi.org/10.1080/0163853X.2015.1023965

Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310).

Gnevsheva, K., Szakay, A., & Jansen, S. (2021). Phonetic convergence across dialect boundaries in first and second language speakers. *Journal of Phonetics*, *89*, 101110. https://doi.org/10.1016/j.wocn.2021.101110

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System*, *60*, 117–127. https://doi.org/10.1016/j.system.2016.06.007

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629. https://doi.org/10.3758/s13414-011-0259-7

Hashemi, M. (2011). Language Stress And Anxiety Among The English Language Learners. *Procedia - Social and Behavioral Sciences*, *30*, 1811–1816. https://doi.org/10.1016/j.sbspro.2011.10.349

Huber, A., & Arthur, H. (2016). Recognizing, Understanding and Mitigating L2 Learner Frustration. *Hirao School of Management review* 6, 170-177. https://doi.org/10.14990/00001738

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72–90. https://doi.org/10.1016/j.jml.2015.01.001

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Kim, I. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *Educational Technology & Society*, *9*, 322–334.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004

Kootstra, G. J., van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language*, *63*(2), 210–231. https://doi.org/10.1016/j.jml.2010.03.006

Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates. http://site.ebrary.com/id/10167345

Kouwenhoven, H., Ernestus, M., & Mulken, M. van. (2018). Register variation by Spanish users of English: The Nijmegen Corpus of Spanish English. *Corpus Linguistics and Linguistic Theory*, *14*(1), 35–63. https://doi.org/10.1515/cllt-2013-0054

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models*. Journal of Statistical Software,* 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Lee, A. H., & Lyster, R. (2017). Can corrective feedback on second language speech perception errors affect production accuracy? *Applied Psycholinguistics*, *38*(2), 371–393. https://doi.org/10.1017/S0142716416000254

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies* (pp. 11-19).

Levitan, R., & Hirschberg, J. (2011). Measuring Acoustic-Prosodic Entrainment with Respect to Multiple Levels and Dimensions. *Interspeech*.

Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *144*(2), 620–633. https://doi.org/10.1121/1.5038567

Li, M., Han, M., Chen, Z., Mo, Y., Chen, X., & Liu, X. (2017). Improving English

Pronunciation Via Automatic Speech Recognition Technology. *2017 International Symposium on Educational Technology (ISET)*, 224–228. https://doi.org/10.1109/ISET.2017.58

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Liu, Q. T. (2017). Phonetic Accommodation to Non-Native English Speech. *UC Berkeley Phonetics and Phonology Lab Annual Report,* spring 2017, p. 108-140.

Loewen, S., & Philp, J. (2006). Recasts in the Adult English L2 Classroom: Characteristics, Explicitness, and Effectiveness. *The Modern Language Journal*, *90*(4), 536–556. https://doi.org/10.1111/j.1540-4781.2006.00465.x

Lombard, E. (1911). Le signe de l’elevation de la voix. *Annales Des Maladies de l’Oreille, Du Larynx, Du Nez et Du Pharynx*, *37*, 101–119.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*(1), 51–81. https://doi.org/10.1017/S027226319800103X

MacIntyre, P. D., & Gardner, R. C. (1991). Language Anxiety: Its Relationship to Other Anxieties and to Processing in Native and Second Languages. *Language Learning*, *41*(4), 513–534. https://doi.org/10.1111/j.1467-1770.1991.tb00691.x

Mackey, A., Al-Khalil, M., Atanassova, G., Hama, M., Logan-Terry, A., & Nakatsukasa, K. (2007). Teachers’ Intentions and Learners’ Perceptions about Corrective Feedback in the L2 Classroom. *Innovation in Language Learning and Teaching*, *1*(1), 129–152. https://doi.org/10.2167/illt047.0

Mackey, A., & Gass, S. (2006). Introduction. *Studies in Second Language Acquisition*, *28*(02). https://doi.org/10.1017/S0272263106060086

MacLeod, B. (2014). Investigating the effects of perceptual salience and regional dialect on phonetic accommodation in Spanish. In M.-H. Côté & E. Mathieu (Eds.), *Current Issues in Linguistic Theory* (Vol. 333, pp. 351–378) https://doi.org/10.1075/cilt.333.24mac

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of Interspeech*, 498–502.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to

EFL language skills. *International Review of Applied Linguistics in Language Teaching*, *56*(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Moore, R. K. (2017b). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In *Dialogues with Social Robots* (pp. 281–291). https://doi.org/10.1007/978-981-10-2585-3_22.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender Differences in Vocal Accommodation: The Role of Perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nassaji, H. (2009). Effects of Recasts and Elicitations in Dyadic Interaction and the Role of Feedback Explicitness. *Language Learning*, *59*(2), 411–452. https://doi.org/10.1111/j.1467-9922.2009.00511.x

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Nushi, M., & Eqbali, M. H. (2017). Duolingo: A Mobile Application to Assist Second Language Learning. *Teaching English with Technology*, 17(1), 89-98.

O'Rourke, E., & Potowski, K. (2016). Phonetic accommodation in a situation of Spanish dialect contact: Coda /s/ and /r̄/ in Chicago. *Studies in Hispanic and Lusophone Linguistics*, *9*(2), 355–399. https://doi.org/10.1515/shll-2016-0015

Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. *Applied Linguistics*, *24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5206–5210. https://doi.org/10.1109/ICASSP.2015.7178964

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190–197. https://doi.

org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254–2264. https://doi.org/10.3758/BF03196699

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183–195. https://doi.org/10.1016/j.jml.2013.06.002

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*(2), 637–659. https://doi.org/10.3758/s13414-016-1226-0

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02). https://doi.org/10.1017/S0140525X04000056

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11). https://doi.org/10.1111/cogs.12911

Raveh, E., Siegert, I., Steiner, I., Gessinger, I., & Möbius, B. (2019). Three's a Crowd? Effects of a Second Human on Vocal Accommodation with a Voice Assistant. *Interspeech 2019*, 4005–4009. https://doi.org/10.21437/Interspeech.2019-1825

Raveh, E., Steiner, I., Siegert, I., Gessinger, I., & Möbius, B. (2019). Comparing phonetic changes in computer-directed and human-directed speech. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung* 2019, 42-49.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. https://doi.org/10.1016/j.jml.2014.05.008

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych

Riek, L. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New

Reporting Guidelines. *Journal of Human-Robot Interaction*, 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

Rojczyk, A. (2012). Phonetic imitation of L2 vowels in a rapid shadowing task. In *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 66-76).

Romera, M., & Elordieta, G. (2013). Prosodic accommodation in language contact: Spanish intonation in Majorca. *International Journal of the Sociology of Language*, *2013*(221). https://doi.org/10.1515/ijsl-2013-0026

Saito, K., & Lyster, R. (2012). Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /ɹ/ by Japanese Learners of English: FFI and L2 Pronunciation. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Schweitzer, K., Walsh, M., & Schweitzer, A. (2017). To See or not to See: Interlocutor Visibility and Likeability Influence Convergence in Intonation. *Interspeech 2017*, 919–923. https://doi.org/10.21437/Interspeech.2017-1248

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Routledge. https://doi.org/10.4324/9780203851357

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, *43*, 26–37. https://doi.org/10.1016/j.wocn.2014.01.004

Skerry-Ryan, R. J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., Rob, C. & Saurous, R. A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. In *International Conference on Machine Learning* (pp. 4693-4702).

Stanley, J. (2018). *joeyr: Functions for Vowel Data*. https://joeystanley.github.io/joeyr/

Staum Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 127-132).

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 16, 163-178.

Tahon, M., Lecorve, G., & Lolive, D. (2020). Can We Generate Emotional Pronunciations for Expressive Speech Synthesis? *IEEE Transactions* *on Affective Computing*, *11*(4), 684–695. https://doi.org/10.1109/TAFFC.2018.2828429

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1105–1122. https://doi.org/10.1098/rstb.2007.2162

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America*, *144*(1), 92–103. https://doi.org/10.1121/1.5044415

Torreira, F. & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica* 69, 124-148. https://doi.org/10.1159/000343635

Troncoso-Ruiz, A., & Elordieta, G. (2018). Prosodic accommodation and salience: The nuclear contours of Andalusian Spanish speakers in Asturias. *Loquens*, *4*(2), 043. https://doi.org/10.3989/loquens.2017.043

Troncoso-Ruiz, A., Ernestus, M. & Broersma, M. (2019). Learning to produce difficult L2 vowels: the effects of awareness-rasing, exposure and feedback. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia 2019, 1094-1098. doi:10.5007/2175-8026.2018v71n3p99.

Tucker, B. V., & Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon*, *11*(3), 375–400. https://doi.org/10.1075/ml.11.3.03tuc

Turnbull, R. (2015). Patterns of individual differences in reduction: Implications for listener-oriented theories. In *Proceedings of the 18th International Congress of Phonetic Sciences*. *International Phonetics Association.*

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech*, *53*(4), 510–540. https://doi.org/10.1177/0023830910372495

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Van Nuffelen, G., Middag, C., De Bodt, M., & Martens, J. (2009). Speech technology-based assessment of phoneme intelligibility in dysarthria. *International Journal of Language & Communication Disorders*, *44*(5), 716–730. https://doi.org/10.1080/13682820802342062

Wang, H., & van Heuven, V. J. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. *Linguistics in the Netherlands*, *23*, 237–248. https://doi.org/10.1075/avt.23.23wan

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*(5), 307–319. https://doi.org/10.1016/j.wocn.2013.03.005

Weide, R. L. (1994). *The CMU Pronouncing Dictionary*. At http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Weisberg, S., & Fox, J. (2011). *An R Companion to Applied Regression*. (2 ed.) Thousand Oaks: Sage.

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, *29*(2–3), 477–491. https://doi.org/10.1016/j.langsci.2006.12.029

Willems, R. M. (2015). *Cognitive Neuroscience of Natural Language Use*. Cambridge University Press.

Willemyns, M., Gallois, C., Callan, V. J., & Pittam, J. (1997). Accent Accommodation in the Job Interview: Impact of Interviewer Accent and Gender. *Journal of Language and Social Psychology*, *16*(1), 3–22. https://doi.org/10.1177/0261927X970161001

Ye, L. (2020). *Do we learn from our mistakes? The effects of communication disruptions on the production-perception link in L2 sound learning*. [Unpublished thesis] Radboud University Nijmegen.

Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age- and Gender-Related Differences in Speech Alignment Toward Humans and Voice-AI. *Frontiers in Communication*, *5*, 600361. https://doi.org/10.3389/fcomm.2020.600361

References

Zhang, F., & Yin, P. (2009). A Study of Pronunciation Problems of English Learners in China. *Asian Social Science*, *5*(6), p141. https://doi.org/10.5539/ass.v5n6p141

Appendix

1. List of the words used, including the critical contrast.

/æ/ word	/ɛ/ word	/æ/ word	/ɛ/ word
Axe	Ex	Land	Lend
Bad	Bed	Man	Men
Bag	Beg	Marry	Merry
Band	Bend	Mat	Met
Bat	Bet	Pan	Pen
Cattle	Kettle	Pat	Pet
Dad	Dead	Sad	Said
Flash	Flesh	Sand	Send
Gas	Guess	Sat	Set
Jam	Gem	Shall	Shell
Had	Head	Tan	Ten
Lad	Led	Than	Then

2. Output from the model fit to friendliness rated with a 5-point scale as predicted by Setting, Exposure and Feedback.

	Friendly		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept: 3|4)	**0.02**	**0.00 – 0.10**	**<0.001**
(Intercept: 4|5)	0.37	0.10 – 1.36	0.134
Setting(Computer)	0.54	0.10 – 3.06	0.489
Condition Factor 2	0.54	0.10 – 3.06	0.489
Condition Factor 3	0.40	0.07 – 2.21	0.294
Condition Factor 4	0.34	0.06 – 1.94	0.226
Setting(Computer):ConditionFactor2	1.60	0.15 – 16.98	0.696
Setting(Computer):ConditionFactor3	5.15	0.46 – 58.08	0.185
Setting(Computer):ConditionFactor4	1.23	0.12 – 12.53	0.859

3. Output from the model fit to the ratings on how well participants thought the other player understood using a 5-point scale as predicted by Setting, Exposure and Feedback.

	Understanding		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept: 2|3)	**0.02**	**0.00 – 0.13**	**<0.001**
(Intercept: 3|4)	0.32	0.10 – 1.07	0.064
(Intercept: 4|5)	**11.71**	**3.06 – 44.78**	**<0.001**
Setting(Computer)	1.95	0.36 – 10.66	0.440
Condition2	1.01	0.19 – 5.41	0.993
Condition3	0.71	0.14 – 3.72	0.689
Condition4	0.78	0.15 – 4.00	0.768
Setting(Computer):Condition2	0.34	0.03 – 3.77	0.380
Setting(Computer):Condition3	1.94	0.18 – 20.84	0.584
Setting(Computer):Condition4	0.60	0.06 – 6.40	0.673

4. Output of the models fit to the F1 and F2 data from the Control condition and the human-human setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F1 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.85**	**0.06**	**45.30**	**13.803**	**<0.001**
Test(Post)	0.04	0.04	22.95	1.185	0.248
Vowel(/ɛ/)	**-0.27**	**0.07**	**277.29**	**-4.016**	**<0.001**
Test(Post): Vowel(/ɛ/)	-0.07	0.05	3406.82	-1.456	0.145
	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.26**	**0.032345**	**89.33**	**8.089**	**<0.001**
Test(Post)	-0.02	0.035691	12.77	-0.444	0.664
Vowel(/ɛ/)	0.006631	0.038500	264.92	0.172	0.863
Test(Post): Vowel(/ɛ/)	0.012251	0.026126	3444.57	0.469	0.639

5. Output of the models fit to the F1 and F2 data from the Control condition and the human-computer setting with the intercept representing the pre-test /æ/ tokens.

Lobanov-transformed F1 values					
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**-0.28**	**0.10**	**12.92**	**-2.673**	**0.019**
Test(Post)	**0.28**	**0.13**	**13.27**	**2.214**	**0.0448**
Vowel(/ε/)	**0.27**	**0.04**	**152.58**	**6.363**	**<0.001**
Test(Post): Vowel(/ε/)	**-0.26**	**0.06**	**4270.86**	**-4.299**	**<0.001**
Lobanov-transformed F2 values					
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.16	0.12	12.82	1.339	0.204
Test(Post)	-0.18	0.17	12.81	-1.030	0.322
Vowel(/ε/)	**-0.13**	**0.04**	**108.68**	**-2.991**	**0.003**
Test(Post): Vowel(/ε/)	**0.27**	**0.06**	**4269.11**	**4.363**	**<0.001**

6. Output of the models fit to the F1 and F2 data from the Control condition and the human-human setting with the intercept representing the post-test /æ/ tokens.

Lobanov-transformed F1 values					
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.00	0.11	12.98	0.032	0.975
Test(Pre)	**-0.28**	**0.13**	**13.37**	**-2.214**	**0.045**
Vowel(/ε/)	0.01	0.04	190.95	0.278	0.781
Test(Pre): Vowel(/ε/)	**0.26**	**0.06**	**4270.86**	**4.299**	**<0.001**
Lobanov-transformed F2 values					
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.02	0.08	14.43	-0.216	0.832
Test(Pre)	0.18	0.17	12.81	1.030	0.322
Vowel(/ε/)	**0.14**	**0.04**	**127.97**	**3.090**	**0.002**
Test(Pre): Vowel(/ε/)	**-0.27**	**0.06**	**4269.11**	**-4.363**	**<0.001**

7. Output of the models fit to the F1 and F2 data from the Control condition and the human-human setting with the intercept representing the pre-test /ɛ/ tokens.

	Lobanov-transformed F1 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.01	0.10	13.02	-0.127	0.901
Test(Post)	0.03	0.13	13.44	0.221	0.828
Vowel(/æ/)	**-0.27**	**0.04**	**152.52**	**-6.363**	**<0.001**
Test(Post): Vowel(/æ/)	**0.26**	**0.06**	**4270.86**	**4.299**	**<0.001**
	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.03	0.12	12.85	0.227	0.824
Test(Post)	0.10	0.17	12.85	0.550	0.592
Vowel(/æ/)	**0.13**	**0.04**	**108.68**	**2.991**	**0.003**
Test(Post): Vowel(/æ/)	**-0.27**	**0.06**	**4269.11**	**-4.363**	**<0.001**

8. Output of the model fit to the F2 data from the OnlyExposure condition and the human-human setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.21**	**0.05**	**21.18**	**4.137**	**<0.001**
Test(Post)	**-0.11**	**0.04**	**13.82**	**-2.557**	**0.023**
Vowel(/ɛ/)	-0.03	0.04	274.49	-0.713	0.477
Test(Post): Vowel(/ɛ/)	**0.23**	**0.03**	**4155.80**	**7.811**	**<0.001**

9. Output of the model fit to the F2 data from the OnlyExposure condition and the human-human setting with the intercept representing the post-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.10**	**0.04**	**40.25**	**2.589**	**0.013**
Test(Pre)	**0.11**	**0.04**	**13.82**	**2.557**	**0.023**
Vowel(/ε/)	**0.20**	**0.04**	**271.96**	**5.255**	**<0.001**
Test(Pre): Vowel(/ε/)	**-0.23**	**0.03**	**4155.80**	**-7.811**	**<0.001**

10. Output of the model fit to the F2 data from the OnlyExposure condition and the human-computer setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.10	0.06	12.02	1.623	0.130
Test(Post)	-0.00	0.11	11.93	-0.012	0.991
Vowel(/ε/)	-0.07	0.04	3596.37	-1.523	0.128
Test(Post): Vowel(/ε/)	**-0.18**	**0.06**	**3595.38**	**-3.038**	**0.002**

11. Output of the model fit to the F2 data from the OnlyExposure condition and the human-computer setting with the intercept representing the post-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.09	0.10	11.87	0.933	0.369
Test(Post)	0.00	0.11	11.93	0.012	0.991
Vowel(/ε/)	**-0.25**	**0.04**	**3594.22**	**-6.192**	**<0.001**
Test(Post): Vowel(/ε/)	**0.18**	**0.06**	**3595.38**	**3.038**	**0.0024**

12. Output of the model fit to the F2 data from the OnlyExposure condition and the human-computer setting with the intercept representing the pre-test /ɛ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.03	0.06	12.45	0.477	0.642
Test(Post)	-0.18	0.11	12.10	-1.624	0.130
Vowel(/æ/)	**0.07**	**0.04**	**3596.37**	**1.523**	**0.128**
Test(Post): Vowel(/æ/)	**0.18**	**0.06**	**3595.38**	**3.038**	**0.002**

13. Output of the model fit to the F1 data from the OnlyFeedback condition with the intercept representing the pre-test /æ/ tokens in the human-human setting.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.84**	**0.07**	**31.27**	**11.816**	**<0.001**
Test(Post)	0.00	0.11	25.74	0.043	0.966
Vowel(/ɛ/)	**-0.22**	**0.04**	**647.22**	**-5.040**	**<0.001**
Setting(Computer)	**-0.65**	**0.10**	**26.55**	**-6.617**	**<0.001**
Test(Post):Vowel(/ɛ/)	-0.12	0.05	8482.34	-2.456	0.014
Test(Post):Setting(Computer)	-0.05	0.15	25.84	-0.344	0.734
Vowel(/ɛ/):Setting(Computer)	**0.14**	**0.05**	**8425.07**	**2.747**	**0.006**
Test(Post):Vowel(/ɛ/):Setting(Computer)	0.01	0.07	8407.49	0.123	0.902

14. Output of the model fit to the F1 data from the OnlyFeedback condition with the intercept representing the post-test /æ/ tokens in the human-human setting.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.84**	**0.07**	**31.33**	**11.618**	**<0.001**
Test(Pre)	-0.00	0.11	25.74	-0.043	0.965
Vowel(/ɛ/)	**-0.35**	**0.04**	**661.44**	**-7.802**	**<0.001**
Setting(Computer)	**-0.70**	**0.10**	**26.91**	**-6.977**	**<0.001**
Test(Pre):Vowel(/ɛ/)	0.12	0.05	8482.34	2.456	0.014
Test(Pre):Setting(Computer)	0.05	0.15	25.84	0.344	0.734
Vowel(/ɛ/):Setting(Computer)	0.15	0.05	8429.76	2.887	0.003
Test(Pre):Vowel(/ɛ/):Setting(Computer)	-0.01	0.07	8407.49	-0.123	0.902

15. Output of the model fit to the F1 data from the OnlyFeedback condition with the intercept representing the post-test /ɛ/ tokens in the human-human setting.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.61**	**0.07**	**29.66**	**8.793**	**<0.001**
Test(Post)	-0.12	0.11	25.96	-1.130	0.269
Vowel(/æ/)	**0.22**	**0.04**	**647.22**	**5.040**	**<0.001**
Setting(Computer)	**-0.50**	**0.10**	**26.79**	**-5.143**	**<0.001**
Test(Post): Vowel(/æ/)	**0.12**	**0.05**	**8482.34**	**2.456**	**0.014**
Test(Post):Setting(Computer)	-0.04	0.15	25.99	-0.284	0.779
Vowel(/æ/):Setting(Computer)	-0.14	0.05	8425.07	-2.747	0.006
Test(Post): Vowel(/æ/):Setting(Computer)	-0.01	0.07	8407.49	-0.123	0.902

16. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-human setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.24**	**0.05**	**20.96**	**4.612**	**<0.001**
Test(Pre)	**-0.11**	**0.034**	**18.67**	**-3.433**	**0.003**
Vowel(/ɛ/)	-0.03	0.04	260.14	-0.836	0.404
Test(Pre): Vowel(/ɛ/)	**0.23**	**0.03**	**4467.18**	**8.385**	**<0.001**

17. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-human setting with the intercept representing the post-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.13	0.04	38.50	3.352	0.002
Test(Pre)	0.11	0.03	18.67	3.433	0.003
Vowel(/ɛ/)	**0.20**	**0.04**	**261.88**	**5.373**	**<0.001**
Test(Pre): Vowel(/ɛ/)	**-0.23**	**0.03**	**4467.18**	**-8.385**	**<0.001**

18. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-human setting with the intercept representing the pre-test /ɛ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.21**	**0.05**	**18.45**	**4.171**	**<0.001**
Test(Post)	**0.11**	**0.03**	**19.36**	**3.282**	**0.004**
Vowel(/æ/)	0.03	0.04	260.14	0.836	0.404
Test(Post): Vowel(/æ/)	**-0.23**	**0.03**	**4467.18**	**-8.385**	**<0.001**

19. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-computer setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**-0.32**	**0.04**	**24.66**	**-8.065**	**<0.001**
Test(Post)	**0.26**	**0.10**	**13.66**	**2.749**	**0.016**
Vowel(/ɛ/)	**0.38**	**0.05**	**188.25**	**8.180**	**<0.001**
Test(Post): Vowel(/ɛ/)	**-0.35**	**0.06**	**3912.88**	**-5.495**	**<0.001**

20. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-computer setting with the intercept representing the post-test /æ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.05	0.08	13.18	-0.632	0.538
Test(Pre)	-0.26	0.10	13.66	-2.749	0.016
Vowel(/ɛ/)	0.03	0.05	192.40	0.718	0.473
Test(Pre): Vowel(/ɛ/)	**0.35**	**0.06**	**3912.88**	**5.495**	**<0.001**

21. Output of the model fit to the F2 data from the OnlyFeedback condition in the human-computer setting with the intercept representing the pre-test /ɛ/ tokens.

	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.06	0.04	25.78	1.600	0.122
Test(Post)	-0.08	0.10	13.77	-0.847	0.412
Vowel(/æ/)	**-0.38**	**0.05**	**188.25**	**-8.180**	**<0.001**
Test(Post): Vowel(/æ/)	**0.35**	**0.06**	**3912.88**	**5.495**	**<0.001**

22. Output of the models fit to the F1 and F2 data from the ExposureAndFeedback condition in the human-human setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F1 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.86**	**0.05**	**109.08**	**17.710**	**<0.001**
Test(Post)	0.02	0.03	22.38	0.606	0.551
Vowel(/ɛ/)	**-0.27**	**0.06**	**299.10**	**-4.684**	**<0.001**
Test(Post): Vowel(/ɛ/)	0.01	0.04	3769.18	0.233	0.816
	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**0.26**	**0.04**	**36.97**	**7.287**	**<0.001**
Test(Post)	-0.04	0.03	13.45	-1.469	0.165
Vowel(/ɛ/)	-0.01	0.03	233.97	-0.250	0.803
Test(Post): Vowel(/ɛ/)	0.03	0.02	3694.69	1.612	0.107

23. Output of the models fit to the F1 and F2 data from the ExposureAndFeedback condition in the human-computer setting with the intercept representing the pre-test /æ/ tokens.

	Lobanov-transformed F1 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.19	0.08	12.96	2.461	0.029
Test(Post)	-0.28	0.17	11.92	-1.566	0.143
Vowel(/ɛ/)	**-0.24**	**0.04**	**3515.07**	**-5.488**	**<0.001**
Test(Post): Vowel(/ɛ/)	**0.23**	**0.06**	**3514.46**	**3.763**	**<0.001**
	Lobanov-transformed F2 values				
	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.05	0.05	19.70	0.966	0.346
Test(Post)	-0.12	0.09	13.40	-1.265	0.227
Vowel(/ɛ/)	-0.08	0.05	501.83	-1.719	0.086
Test(Post): Vowel(/ɛ/)	**0.25**	**0.06**	**3512.66**	**4.166**	**<0.001**

24. Output of the models fit to the F1 and F2 data from the ExposureAndFeedback condition in the human-computer setting with the intercept representing the post-test /æ/.

Lobanov-transformed F1 values	*Estimate*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.08	0.11	11.80	-0.712	0.491
Test(Pre)	0.27	0.17	11.92	1.566	0.143
Vowel(/ɛ/)	-0.02	0.04	3511.92	-0.433	0.665
Test(Pre): Vowel(/ɛ/)	**-0.23**	**0.06**	**3514.46**	**-3.763**	**<0.001**

Lobanov-transformed F2 values	*Estimate*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.07	0.09	13.15	-0.785	0.446
Test(Pre)	0.12	0.09	13.40	1.265	0.227
Vowel(/ɛ/)	**0.17**	**0.04**	**392.96**	**3.707**	**<0.001**
Test(Pre): Vowel(/ɛ/)	**-0.25**	**0.06**	**3512.66**	**-4.166**	**<0.001**

25. Output of the models fit to the F1 and F2 data from the ExposureAndFeedback condition in the human-computer setting with the intercept representing the pre-test /ɛ/.

Lobanov-transformed F1 values	*Estimate*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	-0.10	0.11	11.75	-0.872	0.400
Test(Post)	0.04	0.17	11.90	0.228	0.823
Vowel(/æ/)	0.02	0.04	3511.92	0.433	0.665
Test(Post): Vowel(/æ/)	**0.23**	**0.06**	**3514.46**	**3.763**	**<0.001**

Lobanov-transformed F2 values	*Estimate*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	0.10	0.09	12.76	1.109	0.288
Test(Post)	-0.13	0.09	13.34	-1.467	0.166
Vowel(/æ/)	**-0.17**	**0.04**	**392.96**	**-3.707**	**<0.001**
Test(Post): Vowel(/æ/)	**0.25**	**0.06**	**3512.66**	**4.166**	**<0.001**

26. Output of the logistic regression applied to the human intelligibility data releveled to represent the OnlyExposure condition on the intercept

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	0.24	0.30	0.778	0.436
Test(Post)	**0.22**	**0.06**	**3.633**	**<0.001**
Conditon(Control)	-0.48	0.32	-1.494	0.135
Condition(OnlyFeedback)	**-0.77**	**0.31**	**-2.505**	**0.012**
Condition(Combined)	-0.44	0.32	-1.380	0.167
Test(Post): Conditon(Control)	**-0.27**	**0.09**	**-3.106**	**0.002**
Test(Post): Condition(OnlyFeedback)	0.09	0.09	1.025	0.305
Test(Post): Condition(Combined)	**0.36**	**0.09**	**4.107**	**<0.001**

27. Output of the ordinal regression fitted to the nativelikeness data releveled to represent the Combined condition on the intercept

Predictors	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Test(Post)	-0.01	0.05	-0.119	0.906
Condition(OnlyExposure)	-0.51	0.23	-2.208	0.027
Condition(OnlyFeedback)	0.27	0.22	1.225	0.221
Condition(ExposureAndFeedback)	-0.27	0.22	-1.217	0.224
Test(Post): Condition(OnlyExposure)	**0.26**	**0.07**	**3.469**	**0.001**
Test(Post): Condition(OnlyFeedback)	0.13	0.07	1.764	0.078
Test(Post): Condition(ExposureAndFeedback)	**0.20**	**0.07**	**2.766**	**0.006**

28. Output of the regression fitted to the distance data.

	Estimate	*Std. Error*	*df*	*t value*	*P-value*
(Intercept)	**215.32**	**20.21**	**90.67**	**10.65**	**<0.001**
Test(Post)	-22.63	16.81	86.17	-1.35	0.182
PC1	**-37.10**	**9.77**	**79.72**	**-3.80**	**<0.001**
Condition(OnlyExposure)	-21.60	25.91	80.68	-0.83	0.407
Condition(Combined)	**-64.77**	**26.42**	**80.59**	**-2.45**	**0.016**
Condition(OnlyFeedback)	-31.40	26.26	80.56	-1.20	0.235
Setting(Computer)	**-67.87**	**25.82**	**80.48**	**-2.63**	**0.010**
Test(Post):PC1	**30.45**	**8.14**	**78.85**	**3.74**	**<0.001**
Test(Post):Condition(OnlyExposure)	38.00	21.70	81.47	1.75	0.084
Test(Post):Condition(Combined)	24.88	22.12	81.20	1.12	0.264
Test(Post):Condition(OnlyFeedback)	24.86	21.98	81.16	1.13	0.261
Test(Post):Setting(Computer)	25.92	21.61	80.92	1.20	0.234
PC1:Condition(OnlyExposure)	**39.35**	**15.98**	**80.50**	**2.46**	**0.016**
PC1:Condition(Combined)	**33.88**	**12.70**	**79.77**	**2.67**	**0.009**
PC1:Condition(OnlyFeedback)	**38.99**	**13.09**	**79.92**	**2.98**	**0.004**
PC1:Setting(Computer)	**34.60**	**13.17**	**79.66**	**2.63**	**0.010**
Condition(OnlyExposure):Setting(Computer)	41.84	36.31	80.24	1.15	0.253
Condition(Combined):Setting(Computer)	**91.24**	**37.08**	**80.22**	**2.46**	**0.016**
Condition(OnlyFeedback):Setting(Computer)	35.43	36.69	80.06	0.97	0.337
Test(Post):PC1:Condition(OnlyExposure)	**-41.60**	**13.37**	**80.93**	**-3.11**	**0.003**
Test(Post):PC1:Condition(Combined)	**-30.78**	**10.58**	**78.99**	**-2.91**	**0.005**
Test(Post):PC1:Condition(OnlyFeedback)	**-31.40**	**10.92**	**79.40**	**-2.87**	**0.005**
Test(Post):PC1:Setting(Computer)	**-29.18**	**10.97**	**78.69**	**-2.66**	**0.009**
Test(Post):Condition(OnlyExposure):Setting(Computer)	-28.38	30.34	80.25	-0.93	0.352
Test(Post):Condition(Combined):Setting(Computer)	-6.97	30.98	80.19	-0.22	0.823
Test(Post):Condition(OnlyFeedback):Setting(Computer)	2.04	30.63	79.76	0.07	0.947
PC1:Condition(OnlyExposure):Setting(Computer)	**-48.48**	**20.36**	**80.06**	**-2.38**	**0.020**
PC1:Condition(Combined):Setting(Computer)	-9.56	20.83	79.50	-0.46	0.644
PC1:Condition(OnlyFeedback):Setting(Computer)	**-45.98**	**18.11**	**79.65**	**-2.54**	**0.013**
Test(Post):PC1:Condition(OnlyExposure):Setting(Computer)	**42.56**	**17.00**	**79.75**	**2.50**	**0.014**
Test(Post):PC1:Condition(Combined):Setting(Computer)	2.32	17.33	78.22	0.13	0.894
Test(Post):PC1:Condition(OnlyFeedback):Setting(Computer)	21.30	15.08	78.63	1.41	0.162

29. Output of the model fitted to the vowel distance data relevelled to include the OnlyExposure condition on the intercept.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**193.71**	**18.05**	**91.78**	**10.729**	**<0.001**
Test(Post)	15.37	14.97	84.60	1.026	0.308
PC1	2.25	12.65	80.97	0.178	0.859
Condition(Control)	21.60	25.91	80.68	0.834	0.407
Condition(Combined)	-43.17	24.81	80.46	-1.740	0.086
Condition(OnlyFeedback)	-9.80	24.64	80.43	-0.398	0.692
Setting(Computer)	-26.03	25.53	80.00	-1.019	0.311
Test(Post):PC1	-11.15	10.61	82.18	-1.051	0.296
Test(Post):Condition(Control)	-37.99	21.70	81.46	-1.751	0.084
Test(Post):Condition(Combined)	-13.12	20.76	80.83	-0.632	0.529
Test(Post):Condition(OnlyFeedback)	-13.14	20.61	80.77	-0.637	0.526
Test(Post):Setting(Computer)	-2.45	21.30	79.58	-0.115	0.909
PC1:Condition(Control)	**-39.35**	**15.98**	**80.50**	**-2.462**	**0.016**
PC1:Condition(Combined)	-5.47	15.03	80.64	-0.364	0.717
PC1:Condition(OnlyFeedback)	-0.37	15.36	80.72	-0.024	0.981
PC1:Setting(Computer)	-13.88	15.53	80.35	-0.894	0.374
Condition(Control):Setting(Computer)	-41.84	36.31	80.24	-1.152	0.253
Condition(Combined):Setting(Computer)	49.39	36.88	79.99	1.339	0.184
Condition(OnlyFeedback):Setting(Computer)	-6.41	36.49	79.82	-0.176	0.861
Test(Post):PC1:Condition(Control)	**41.60**	**13.37**	**80.93**	**3.110**	**0.003**
Test(Post):PC1:Condition(Combined)	10.82	12.59	81.30	0.860	0.392
Test(Post):PC1:Condition(OnlyFeedback)	10.20	12.87	81.52	0.792	0.431
Test(Post):PC1:Setting(Computer)	13.38	12.98	80.52	1.030	0.306
Test(Post):Condition(Control):Setting(Computer)	28.38	30.34	80.25	0.935	0.352
Test(Post):Condition(Combined):Setting(Computer)	21.41	30.77	79.55	0.696	0.489
Test(Post):Condition(OnlyFeedback):Setting(Computer)	30.42	30.41	79.10	1.000	0.320
PC1:Condition(Control):Setting(Computer)	**48.48**	**20.36**	**80.06**	**2.381**	**0.02**
PC1:Condition(Combined):Setting(Computer)	38.82	22.40	79.85	1.733	0.087
PC1:Condition(OnlyFeedback):Setting(Computer)	2.50	19.89	80.07	0.125	0.900
Test(Post):PC1:Condition(Control):Setting(Computer)	**-42.56**	**17.00**	**79.75**	**-2.504**	**0.014**
Test(Post):PC1:Condition(Combined):Setting(Computer)	**-40.24**	**18.67**	**79.15**	**-2.155**	**0.034**
Test(Post):PC1:Condition(OnlyFeedback):Setting(Comput.)	-21.26	16.61	79.76	-1.281	0.204

30. Output of the model fitted to the vowel distance data relevelled to include the Only Feedback condition on the intercept.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**183.92**	**18.55**	**91.20**	**9.913**	**<0.001**
Test(Post)	2.23	15.38	84.30	0.145	0.885
PC1	1.89	8.72	80.19	0.217	0.829
Condition(OnlyExposure)	9.80	24.64	80.43	0.398	0.692
Condition(Control)	31.40	26.26	80.56	1.196	0.235
Condition(Combined)	-33.38	25.18	80.33	-1.326	0.189
Setting(Computer)	-32.44	26.07	79.64	-1.244	0.217
Test(Post):PC1	-0.95	7.28	80.10	-0.131	0.896
Test(Post):Condition(OnlyExposure)	13.14	20.60	80.77	0.637	0.526
Test(Post):Condition(Control)	-24.86	21.98	81.16	-1.131	0.261
Test(Post):Condition(Combined)	0.02	21.05	80.50	0.001	0.999
Test(Post):Setting(Computer)	27.96	21.71	78.64	1.288	0.201
PC1:Condition(OnlyExposure)	0.37	15.36	80.72	0.024	0.981
PC1:Condition(Control)	**-38.99**	**13.09**	**79.92**	**-2.977**	**0.004**
PC1:Condition(Combined)	-5.10	11.91	80.03	-0.429	0.669
PC1:Setting(Computer)	-11.38	12.43	79.63	-0.916	0.363
Condition(OnlyExposure):Setting(Computer)	6.41	36.48	79.82	0.176	0.861
Condition(Control):Setting(Computer)	-35.43	36.69	80.06	-0.966	0.337
Condition(Combined):Setting(Computer)	55.81	37.25	79.82	1.498	0.138
Test(Post):PC1:Condition(OnlyExposure)	-10.19	12.87	81.52	-0.792	0.431
Test(Post):PC1:Condition(Control)	**31.40**	**10.92**	**79.40**	**2.875**	**0.005**
Test(Post):PC1:Condition(Combined)	0.62	9.94	79.68	0.063	0.950
Test(Post):PC1:Setting(Computer)	-7.88	10.35	78.57	-0.762	0.448
Test(Post):Condition(OnlyExposure):Setting(Computer)	-30.41	30.41	79.10	-1.000	0.32
Test(Post):Condition(Control):Setting(Computer)	-2.04	30.63	79.76	-0.067	0.947
Test(Post):Condition(Combined):Setting(Computer)	-9.01	31.05	79.09	-0.290	0.772
PC1:Condition(OnlyExposure):Setting(Computer)	-2.49	19.89	80.07	-0.125	0.900
PC1:Condition(Control):Setting(Computer)	**45.98**	**18.11**	**79.65**	**2.539**	**0.013**
PC1:Condition(Combined):Setting(Computer)	36.32	20.37	79.47	1.783	0.078
Test(Post):PC1:Condition(OnlyExposure):Setting(Computer)	21.26	16.61	79.75	1.281	0.204
Test(Post):PC1:Condition(Control):Setting(Computer)	-21.30	15.08	78.63	-1.412	0.162
Test(Post):PC1:Condition(Combined):Setting(Computer)	-18.98	16.94	78.15	-1.120	0.266

31. Output of the model fitted to the vowel distance data relevelled to include the Combined condition on the intercept.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**150.54**	**18.79**	**91.12**	**8.014**	**<0.001**
Test(Post)	2.25	15.58	84.52	0.144	0.886
PC1	-3.22	8.11	79.86	-0.396	0.693
Condition(OnlyFeedback)	33.38	25.18	80.33	1.326	0.189
Condition(OnlyExposure)	43.17	24.81	80.46	1.740	0.086
Condition(Control)	**64.77**	**26.42**	**80.59**	**2.451**	**0.016**
Setting(Computer)	23.37	26.61	79.98	0.878	0.383
Test(Post):PC1	-0.33	6.76	79.19	-0.049	0.961
Test(Post):Condition(OnlyFeedback)	-0.02	21.05	80.50	-0.001	0.999
Test(Post):Condition(OnlyExposure)	13.12	20.76	80.83	0.632	0.529
Test(Post):Condition(Control)	-24.88	22.12	81.20	-1.125	0.264
Test(Post):Setting(Computer)	18.96	22.20	79.52	0.854	0.396
PC1:Condition(OnlyFeedback)	5.10	11.91	80.03	0.429	0.669
PC1:Condition(OnlyExposure)	5.47	15.03	80.64	0.364	0.717
PC1:Condition(Control)	**-33.88**	**12.70**	**79.77**	**-2.668**	**0.009**
PC1:Setting(Computer)	24.94	16.14	79.39	1.545	0.126
Condition(OnlyFeedback):Setting(Computer)	-55.81	37.25	79.82	-1.498	0.138
Condition(OnlyExposure):Setting(Computer)	-49.39	36.88	79.99	-1.339	0.184
Condition(Control):Setting(Computer)	-91.24	37.08	80.22	-2.460	0.016
Test(Post):PC1:Condition(OnlyFeedback)	-0.62	9.94	79.67	-0.063	0.95
Test(Post):PC1:Condition(OnlyExposure)	-10.82	12.59	81.30	-0.860	0.392
Test(Post):PC1:Condition(Control)	**30.78**	**10.58**	**78.99**	**2.909**	**0.004**
Test(Post):PC1:Setting(Computer)	-26.86	13.42	77.91	-2.001	0.049
Test(Post):Condition(OnlyFeedback):Setting(Computer)	9.01	31.05	79.08	0.290	0.772
Test(Post):Condition(OnlyExposure):Setting(Computer)	-21.41	30.77	79.55	-0.696	0.489
Test(Post):Condition(Control):Setting(Computer)	6.97	30.98	80.19	0.225	0.822
PC1:Condition(OnlyFeedback):Setting(Computer)	-36.32	20.37	79.47	-1.783	0.078
PC1:Condition(OnlyExposure):Setting(Computer)	-38.82	22.40	79.85	-1.733	0.087
PC1:Condition(Control):Setting(Computer)	9.66	20.83	79.50	0.464	0.644
Test(Post):PC1:Condition(OnlyFeedback):Setting(Comput.)	18.98	16.95	78.15	1.120	0.266
Test(Post):PC1:Condition(OnlyExposure):Setting(Comput.)	**40.24**	**18.67**	**79.15**	**2.155**	**0.034**
Test(Post):PC1:Condition(Control):Setting(Computer)	-2.32	17.33	78.22	-0.134	0.894

32. Output of the model fitted to the vowel distance data relevelled to include the human-computer setting on the intercept.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**147.45**	**17.93**	**91.35**	**8.221**	**<0.001**
Test(Post)	3.29	14.84	83.31	0.222	0.825
PC1	-2.50	8.83	79.60	-0.283	0.778
Condition(Combined)	26.46	26.02	79.85	1.017	0.312
Condition(OnlyFeedback)	4.03	25.63	79.53	0.157	0.875
Condition(OnlyExposure)	20.24	25.45	79.79	0.795	0.428
Setting(Human)	**67.87**	**25.82**	**80.48**	**2.628**	**0.01**
Test(Post):PC1	1.27	7.35	78.51	0.172	0.863
Test(Post):Condition(Combined)	17.91	21.69	79.17	0.826	0.411
Test(Post):Condition(OnlyFeedback)	26.9	21.33	78.31	1.261	0.211
Test(Post):Condition(OnlyExposure)	9.62	21.21	79.01	0.454	0.651
Test(Post):Setting(Human)	-25.92	21.61	80.92	-1.200	0.234
PC1:Condition(Combined)	24.22	16.52	79.33	1.467	0.146
PC1:Condition(OnlyFeedback)	-6.99	12.51	79.34	-0.559	0.578
PC1:Condition(OnlyExposure)	-9.12	12.62	79.38	-0.723	0.472
PC1:Setting(Human)	**-34.60**	**13.17**	**79.66**	**-2.627**	**0.01**
Condition(Combined):Setting(Human)	**-91.24**	**37.08**	**80.22**	**-2.460**	**0.016**
Condition(OnlyFeedback):Setting(Human)	-35.43	36.69	80.06	-0.966	0.337
Condition(OnlyExposure):Setting(Human)	-41.84	36.31	80.24	-1.152	0.253
Test(Post):PC1:Condition(Combined)	**-28.46**	**13.73**	**77.77**	**-2.073**	**0.041**
Test(Post):PC1:Condition(OnlyFeedback)	-10.10	10.40	77.80	-0.971	0.334
Test(Post):PC1:Condition(OnlyExposure)	0.96	10.49	77.90	0.092	0.927
Test(Post):PC1:Setting(Human)	**29.18**	**10.97**	**78.69**	**2.661**	**0.009**
Test(Post):Condition(Combined):Setting(Human)	6.97	30.98	80.19	0.225	0.823
Test(Post):Condition(OnlyFeedback):Setting(Human)	-2.04	30.63	79.76	-0.067	0.947
Test(Post):Condition(OnlyExposure):Setting(Human)	28.38	30.34	80.25	0.935	0.352
PC1:Condition(Combined):Setting(Human)	9.66	20.83	79.50	0.464	0.644
PC1:Condition(OnlyFeedback):Setting(Human)	**45.98**	**18.11**	**79.65**	**2.539**	**0.013**
PC1:Condition(OnlyExposure):Setting(Human)	**48.48**	**20.36**	**80.06**	**2.381**	**0.02**
Test(Post):PC1:Condition(Combined):Setting(Human)	-2.32	17.33	78.22	-0.134	0.894
Test(Post):PC1:Condition(OnlyFeedback):Setting(Human)	-21.30	15.08	78.63	-1.412	0.162
Test(Post):PC1:Condition(OnlyExposure):Setting(Human)	**-42.56**	**17.00**	**79.75**	**-2.504**	**0.014**

33. Output of the model fitted to vowel distance data relevelled to include the Combined condition of the human-computer setting.

	Estimate	*Std. Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	**173.90**	**20.46**	**89.32**	**8.499**	**<0.001**
Test(Post)	21.21	16.92	83.18	1.253	0.214
PC1	21.72	13.95	79.23	1.557	0.123
Condition(Control)	-26.46	26.02	79.85	-1.017	0.312
Condition(OnlyFeedback)	-22.43	27.46	79.39	-0.817	0.416
Condition(OnlyExposure)	-6.22	27.29	79.61	-0.228	0.820
Setting(Human)	-23.37	26.61	79.98	-0.878	0.383
Test(Post):PC1	**-27.19**	**11.59**	**77.48**	**-2.346**	**0.021**
Test(Post):Condition(Control)	-17.91	21.69	79.17	-0.826	0.411
Test(Post):Condition(OnlyFeedback)	8.99	22.83	77.91	0.394	0.695
Test(Post):Condition(OnlyExposure)	-8.29	22.71	78.50	-0.365	0.716
Test(Post):Setting(Human)	-18.96	22.20	79.52	-0.854	0.396
PC1:Condition(Control)	-24.22	16.52	79.33	-1.467	0.146
PC1:Condition(OnlyFeedback)	-31.22	16.53	79.19	-1.889	0.063
PC1:Condition(OnlyExposure)	**-33.35**	**16.61**	**79.21**	**-2.008**	**0.048**
PC1:Setting(Human)	-24.94	16.14	79.39	-1.545	0.126
Condition(Control):Setting(Human)	**91.24**	**37.08**	**80.22**	**2.460**	**0.016**
Condition(OnlyFeedback):Setting(Human)	55.80	37.25	79.82	1.498	0.138
Condition(OnlyExposure):Setting(Human)	49.39	36.88	79.99	1.339	0.184
Test(Post):PC1:Condition(Control)	**28.46**	**13.73**	**77.77**	**2.073**	**0.041**
Test(Post):PC1:Condition(OnlyFeedback)	18.35	13.73	77.37	1.337	0.185
Test(Post):PC1:Condition(OnlyExposure)	**29.42**	**13.79**	**77.43**	**2.133**	**0.036**
Test(Post):PC1:Setting(Human)	**26.86**	**13.42**	**77.91**	**2.001**	**0.049**
Test(Post):Condition(Control):Setting(Human)	-6.97	30.98	80.19	-0.225	0.823
Test(Post):Condition(OnlyFeedback):Setting(Human)	-9.01	31.05	79.08	-0.290	0.772
Test(Post):Condition(OnlyExposure):Setting(Human)	21.41	30.77	79.55	0.696	0.489
PC1:Condition(Control):Setting(Human)	-9.66	20.83	79.50	-0.464	0.644
PC1:Condition(OnlyFeedback):Setting(Human)	36.32	20.37	79.47	1.783	0.078
PC1:Condition(OnlyExposure):Setting(Human)	38.82	22.40	79.85	1.733	0.087
Test(Post):PC1:Condition(Control):Setting(Human)	2.32	17.33	78.22	0.134	0.894
Test(Post):PC1:Condition(OnlyFeedback):Setting(Human)	-18.98	16.95	78.15	-1.120	0.266
Test(Post):PC1:Condition(OnlyExposure):Setting(Human)	**-40.24**	**18.67**	**79.15**	**-2.155**	**0.034**

34. Table with the correlation between the overall nativelikeness score and other proficiency measures, including individual measures and the PC1 for the subset of the participants in the human-human setting.

Measure	*Correlation to the nativelikeness score*
Lextale score	0.11
Proficiency Listening	0.14
Proficiency Writing	0.22
Proficiency Reading	0.07
Proficiency Speaking	0.3
Dutch Accentedness	0.35
Identification as a Non-native Speaker	0.28
PC1	0.39

Research Data Management

Personal data

Will you process personal data? If yes, how will you ensure compliance with legislation on privacy?

Yes, I process personal data during my research.
I will collect or process the following personal data: speech recordings, age, gender and L2 experience. These data contains personal data: speech recordings.
It is necessary to collect or process these personal data to achieve the goals of my project because: we need to analyse speech recordings, we need to ensure the sample population is as homogeneous as possible (age and gender) and the L2 experience is used in one of our research questions (we investigate whether L2 proficiency modulates the behaviour of participants during an interaction).
I will ensure that I will not collect more personal data than necessary for achieving the goals of my research project as follows: effect of proficiency in phonetic adaptation
I am going to retain the personal data for the following time period: until the end of my academic career.

In order to protect the privacy of your participants, will you anonymise or pseudonymise the data?

I will protect the privacy of my participants by anonymising or pseudonymising (some of) the personal data.
I will do this in the following manner: all the personal data will be anonymised upon collection by using numerical identifiers to refer to participants. These identifiers will be used in filenames and datasets. The personal information about the participants (included in the consent form) will not be used. No document will exist linking the personal information included in the consent forms with the participant numeric identifier. The consent forms will be securely stored in the CLS lab.

The content of the speech recordings cannot be anonymised. I asked participants for permission to collect and share their data for research purposes. Only the data of the participants who agreed will be shared.

Do you need approval from an ethics committee for your project?

Yes, for my research I will receive/have received approval from an ethics committee. I received approval from Ethics Assessment Committee of the Faculty of Arts (EAC) on 15.12.2016

Does your research require an informed consent procedure?

Yes, I work with human participant data and need an informed consent procedure. I will follow the informed consent procedure established by the Ethics Assessment Committee Humanities at Radboud University, as specified here.

Storing and sharing during research

Will you make use of safe storage during your research, including back-up facilities?

Yes, I will follow my institute's policy and use the work group folders. I will store my data in a Radboud University work group folder (i.e., "werkgroepmap") on the university's network drive. These folders are automatically backed up on a daily basis. When off-campus, I can securely access the data through a VPN connection. I will make sure that the structure of my Radboud University work group folder meets the minimum requirements as described in my institute's RDM protocol.

With whom will you share your data during research?

During my research I need to share my data with researchers and/or students affiliated with Radboud University. I will follow the policy of my institute and use Surfdrive.
I need to share the data with researchers and/or students affiliated with Radboud University because we collaborate together in research projects related to the data collected.

How will you deal with security issues that arise during your research?

According to our policy, the data are stored in workgroup folders ("werkgroepmappen") while research is ongoing This storage location meets legal and ethical requirements. Safe and secure storage is guaranteed by the IT security and safety protocols.
In addition, Surfdrive or FileSender will be used to exchange standard data between researchers during the project. Please select: Workgroup folders will be used to exchange personal data, or alternatively, FileSender in encryption mode. When I am gathering personal data off-campus, I will use a secure VPN connection to transfer the data to the workgroup folders as soon as possible. Before my data will be transferred I will save my data on an encrypted laptop/external harddrive/USB stick. Because my data is stored on the University's network drives, the data is automatically backed up on a daily basis.

I organise my project's folder according to the following format:

I will make sure that the structure of my research data folder meets the minimum requirements as described in my institute's RDM protocol.

Long term archiving and reuse

In the context of scientific integrity, where will you archive your data (including raw data, metadata and documentation) for at least 10 years?

I will follow the policy of my institute and will make sure to archive the research data associated with my publication (including raw data, metadata and documentation) in a closed collection in the Radboud Data Repository (data.ru.nl) for a minimum of 10 years.

In the context of data reuse, will you make your research data publicly available?

Yes, I will follow my institute's policy and make my data public at publication via the Radboud Data Repository (https://data.ru.nl). This includes metadata (Dublin Core/Datacite), documentation, and one of the available open access licenses.

How will you ensure that your research data will be stored in a FAIR manner?

I will use the Radboud Data Repository to archive my data. My data will comply with the FAIR principles in the following way: where data can be made public, my data will be Findable via a Data Sharing Collection in the Radboud Data Repository, which is indexed by search engines on the internet. My data includes rich metadata and has a persistent identifier (DOI).

My data will be Accessible as well, since the Radboud Data Repository uses an open internet protocol, including clear authorisation procedures.

My data will be Interoperable by the use of standards for metadata (Dublin Core/Datacite), standard preferred data formats and domain-specific vocabularies (if existing).

My data will be Reusable via the Radboud Data Repository, including rich metadata and documentation for reuse, and a clear license.

English Summary

When learning a second language (L2), learners often struggle with the aspects related to the pronunciation. As a result, the speech of non-native speakers is often accented with traces to their native language (L1). For example, many Dutch speakers often pronounce the English consonant /θ/, in words like "**th**ink", as a /d/ or a /t/ (Hanulíková & Weber, 2012; Wester et al., 2007), and the vowel /æ/, in "b**a**d", as an /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006). These pronunciation difficulties often emerge from the differences between the sound existing in the L1 and the L2. In the previous examples, Dutch speaker substitute the English sounds /θ/ and /æ/, not present in Dutch, for others sounds that do exist in Dutch: /d/ or /t/, and /ɛ/, respectively.

One of the ways in which L2 speakers can learn to overcome these struggles is by using the L2 in interactions. When interacting, speakers often adapt the way they speak to the interlocutor (the conversation partner). For example, speakers of a stigmatized dialect tend to change the way they speak when interacting with a speaker of a different (less stigmatized) variety. This phenomenon, known as *accommodation* (Giles & Ogay, 2007), can affect the selection of words (Brennan & Clark, 1996), the syntactic structures (Branigan et al., 2000), and also the pronunciation (Pardo, 2006). Accommodation often results in the interlocutors sounding more similar to each other. Accommodation could be a driving mechanism for the learning of L2 pronunciation to take place: if a non-native speaker interacts with a native speaker, the non-native speaker can accommodate their pronunciation to sound more similar to the native pronunciation of the interlocutor and, consequently, sound more native-like than before the interaction. If these adaptations are preserved after the conversation, accommodation could lead to L2 pronunciation learning.

The origin of accommodation, why people accommodate in the first place, is often the focus of the literature investigating this phenomenon. Researchers often study which social and linguistic factors trigger the presence of accommodation in interactions. For example, a speaker's age (Zellou et al., 2021), the role they have in the conversation (Pardo, 2006), and the power dynamics between the two interlocutors (Willemyns et al., 1997) can influence the degree and direction of accommodation. From the social perspective, accommodation is interpreted as a mechanism to bridge social gaps between the interlocutors (Giles & Ogay, 2007). From a linguistic perspective, this phenomenon has been proposed to be the product of efficiency (Pickering & Garrod, 2004), facilitating the role of the speaker or the role of the interlocutor in the interaction. The current dissertation focuses on investigating the linguistic factors that may trigger non-native speakers to accommodate during an interaction.

Pickering and Garrod (2004) proposed that accommodation is an automatic mechanism that facilitates the role of the speaker. From this perspective, accommodation is the product of priming, i.e., the process by which exposure to an item facilitates processing of that item in the future. If, during a conversation, a speaker receives exposure to a particular element from the interlocutor, for example, the word "sofa"; it would be easier/more efficient for them to use this element again in the interaction, than using other similar elements such as "couch". Therefore, exposure to the interlocutor's speech would lead the speaker to accommodate and use the same communicating patterns the interlocutor uses, for efficiency reasons.

Accommodation has also been proposed to be a mechanism facilitating the listening interlocutor's role in an interaction (Clark, 1996). A speaker can accommodate their speech to the interlocutor in order to ensure that the interlocutor will understand the message. For example, a speaker may choose to use the word "sofa" (instead of "couch") if they think this word is more likely to be understood by the interlocutor. Thus, accommodation would take place in order to ensure understanding and prevent miscommunications.

Previous literature seems to indicate that the weight of these two perspectives heavily depends on the circumstances of the conversation, for example, whether the speaker is using their native language or a second language (Costa et al., 2008) and whether the interlocutor is a person or a computer (Branigan et al., 2010). In order to research the mechanism underlying accommodation, researchers study conversations between two interlocutors in a laboratory setting.

With a few exceptions (Berry & Ernestus, 2018; Romera & Elordieta, 2013), most studies tend to sacrifice the ecological validity of the interaction, i.e., the spontaneity and naturalness of the conversation, in order to exercise exhaustive control over the interaction taking place in the lab. For example, the interlocutor's voice is often pre-recorded and presented to the participants using a computer. As a consequence, the participants in these experiments interact with a computer, instead of with a person. Moreover, the tasks often used in these recreations tend to be extremely simplified versions of interactions, such as, for example, shadowing tasks, where participants have to repeat isolated words played by the computer (Cohn et al., 2019). These circumstances hinder the generalizability to real-life human-human interactions.

This dissertation aims to expand our knowledge about non-native phonetic accommodation by investigating what mechanisms underlie the learning of L2 pronunciation in interactions. The dissertation consists of an introduction, four experimental chapters and a general discussion. The studies described in the experimental chapters investigate phonetic accommodation in Dutch learners of English who participant in an interaction in English. As mentioned above, this group of speakers tend to struggle with the pronunciation of the contrast between the English vowels in "bad" and "bed" (/bæd-bɛd/), since they tend to pronounce both vowels as an intermediate sound more similar to the vowel /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006).

Chapter 2 presents a new experimental paradigm, the Ventriloquist paradigm, developed for the study of conversations in laboratory settings. With this paradigm, participants interact with a confederate, a person who is an accomplice in the experiment. Both the participant and the confederate are sitting in the same room and are equipped with headphones and microphones while playing a co-operative game together. The participant believes that they listen to the confederate's voice through the headphones but, in reality, they listen to pre-recorded samples of speech. When the confederate wants to "speak", she partially hides her face behind a monitor and presses a button on a hidden keyboard to play the pre-recorded audio fragment. This methodology successfully combines the exhaustive control over the phonetic input the participant receives and ecological validity.

This chapter includes a detailed description of the implementation of this paradigm in two different experiments with two different confederates. In addition, this chapter shows that the credibility of the paradigm is above acceptable: around 80% of the participants reported believing that they were talking to the interlocutor directly and did not notice the manipulation. Finally, this chapter also confirms that this paradigm is more efficient in terms of ecological validity than other paradigms commonly used to study accommodation; in particular, those where human confederates are replaced with computers.

Chapter 3 describes a study investigating whether non-native speakers accommodate their English pronunciation differently in interactions with humans and in interactions with computers. In order to do so, two identical settings were created: one where non-native speakers interacted with a human confederate and one where they interacted with a computer. Using the Ventriloquist paradigm and pre-recorded speech, the speech of the confederate, human or computer, was controlled to be the same across settings. In particular, we investigated whether specific factors – such as exposure to the confederate's pronunciation of an English sound and feedback implying misunderstanding from the interlocutor–, triggered phonetic accommodation differently in these two settings. In order to do this, the presence or absence of corrective feedback from the interlocutor was implemented. Pre-recording the interlocutor's speech allows for the strict control over the presence or absence of exposure to certain sounds, such as the vowel in question, in the speech of the interlocutor throughout the conversation. Implicit corrective feedback was also implemented in the experiment. The feedback indicated that the way in which the participant pronounces the English vowel /æ/ caused misunderstandings during the conversation. In particular, the interlocutor pretended to understand words with the vowel /ɛ/, for example, “bed”, whenever the participant said words including the vowel /æ/, for example, “bad”.

The results of this chapter, based on the acoustic analyses of the vowels of interest, show that Dutch speakers who received exposure to the native pronunciation of the vowel /æ/ accommodated more when the exposure came from the human interlocutor than from the computer. Similarly, feedback on the non-native speaker’s pronunciation was more effective in eliciting accommodation when it came from a person than from a computer. Surprisingly, the combination of both factors, exposure and feedback, did not trigger accommodation in the speakers who interacted with the human interlocutor. In the group of participants who interacted with the computer, the combination of the two factors triggered more accommodation than when the two factors were presented in isolation.

Previous literature highlights the importance of using acoustic analyzes in combination with perceptual measures to account for the degree of accommodation. In Chapter 4, three different perceptual measures are used in order to investigate the effects of exposure to the speaker’s pronunciation and corrective feedback on the degree of accommodation in the context of human-to-human interactions. The three perceptual measures include intelligibility and nativelikeness ratings produced by native listeners of English and intelligibility judgements generated using automatic speech alignment software.

The results indicate that the intelligibility increased for those non-native speakers who received exposure to the interlocutor’s pronunciation and feedback on their own pronunciation during the interaction. The improvement was even greater when the factors were combined. Second, results show that the Dutch speakers’ English pronunciation sounded more native-like after the conversation, only when the speakers had received exposure or feedback separately, but not in combination. Furthermore, participants' pronunciation was also evaluated using automatic speech alignment software. These types of programs can offer an efficient alternative to human assessment of speech intelligibility. The results show that the software generated intelligibility evaluations overall similar to those produced by native listeners. However, key differences were found between the two intelligibility measurements.

Chapter 5 discusses the effect of the level of proficiency on phonetic accommodation by second language speakers. Previous literature offers contradictory predictions about whether proficiency enhances or hinders non-native accommodation. On the one hand, since using a second language already requires considerable cognitive effort, more proficient speakers could be expected to have more cognitive resources available to accommodate than less proficient second language speakers. On the other hand, if accommodation is understood as a phenomenon aimed to ensure a successful interaction, speakers with lower proficiency, whose accented pronunciation may pose a greater risk to the success of the interaction, could show more accommodation than high proficiency speakers, who generally sound more native-like.

This chapter offers a re-analysis of the data presented in Chapter 3, taking into account the level of English proficiency of the Dutch speakers. In the human-human setting, proficiency was found to facilitate accommodation: high proficiency speakers accommodated to the human interlocutor more than low proficiency speakers. This effect was only found in the "control" condition of the experiment, where participants did not receive neither exposure nor feedback. In the human-computer setting, on the contrary, proficiency was found to hinder accommodation: low proficiency speakers accommodated more than high proficiency speakers. This effect was only found in the condition where the speakers received both feedback on their pronunciation and exposure to the speaker's pronunciation.

The chapters included in this dissertation document the presence of phonetic accommodation in non-native speakers. In addition, accommodation is found to be both an automatic phenomenon, facilitating the role of the speaker, and an audience-design process, facilitating the task of the listener. Therefore, this dissertation supports a definition that combines both perspectives. In addition, the results of these studies show that the degree of accommodation drastically depends on the type of interlocutor, whether it is a human being or a computer. Accommodation to computers is also documented in these studies, but the key differences found between accommodation to computers and accommodation to humans indicate that computer-based methodologies are not directly generalizable to interactions between humans.

Nederlandse Samenvatting

Bij het leren van een tweede taal (T2) is de uitspraak vaak een van de moeilijkste aspecten om onder de knie te krijgen. In de praktijk bevat het accent van tweedetaalsprekers meestal sporen van hun moedertaal (T1). Veel Nederlandstaligen – zelfs degenen met een hoog T2 taalniveau– worstelen bijvoorbeeld vaak met de uitspraak van bepaalde Engelse klanken. Deze groep tweedetaalsprekers spreekt de Engelse medeklinker /θ/, in woorden als ***th**ink*, vaak uit als een /d/ of een /t/ (Hanulíková & Weber, 2012; Wester et al., 2007), en de klinker /æ/, in *b**a**d*, als /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006). Deze moeilijkheden komen vaak voort uit de verschillen tussen de geluiden die voorkomen in de T1 en de T2. De volgende Engelse klanken, /θ/ en /æ/, bestaan niet in het Nederlands. Daarom hebben Nederlandstaligen de neiging ze te vervangen door andere geluiden die wel in hun T1 bestaan: /d/ of /t/ en /ɛ/ respectievelijk.

Een manier om deze moeilijkheden in de uitspraak van een vreemde taal te overwinnen, is door de T2 in interacties te gebruiken. Tijdens een gesprek hebben sprekers de neiging om de manier van communiceren aan te passen aan de spraak van de gesprekspartner. Sprekers van een gestigmatiseerd dialect hebben bijvoorbeeld de neiging om de manier van spreken te veranderen wanneer ze communiceren met een spreker van een andere (minder gestigmatiseerde) dialect. Dit fenomeen is bekend als accommodatie – *accommodation* in het Engels (Giles & Ogay, 2007), bijvoorbeeld de selectie van gebruikte woorden (Brennan & Clark, 1996), de syntactische structuren (Branigan et al., 2000), en ook de uitspraak (Pardo, 2006). Als gevolg van accommodatie gaat de spraak van sprekers meer op elkaar lijken. Accommodatie kan een drijvende kracht zijn achter het leren van de T2-uitspraak: als een niet-moedertaalspreker met een moedertaalspreker praat, kan de niet-moedertaalspreker haar uitspraak aanpassen om meer op de uitspraak van de gesprekspartner te lijken en, daardoor klinken ze meer meer als de moedertaalspreker dan vóór de interactie. Als deze aanpassingen na het gesprek behouden blijven, kan accommodatie leiden tot het beter leren van de uitspraak van T2. De oorsprong van accommodatie, dus waarom mensen in de eerste plaats accommoderen, is vaak de focus van de literatuur die dit fenomeen onderzoekt. Onderzoekers bestuderen vaak welke sociale en linguïstische factoren de aanwezigheid van accommodatie in interacties veroorzaken. Zo kunnen de leeftijd van een spreker (Zellou et al., 2021), de rol die ze hebben in het gesprek (Pardo, 2006) en de machtsdynamiek tussen de twee gesprekspartners (Willemyns et al., 1997) de mate en richting van accommodatie beïnvloeden. Vanuit sociaal perspectief wordt accommodatie geïnterpreteerd als een mechanisme om de sociale kloof tussen de gesprekspartners te overbruggen (Giles & Ogay, 2007). Vanuit een taalkundig perspectief is voorgesteld dat dit fenomeen het product van efficiëntie is (Pickering & Garrod, 2004), waardoor de rol van de spreker of de rol van de gesprekspartner in de interactie wordt vergemakkelijkt. Het huidige proefschrift richt zich op het onderzoeken van de linguïstische factoren die tweedetaalsprekers ertoe kunnen aanzetten zich tijdens een interactie aan te passen.

Pickering & Garrod (2004) stelden voor dat accommodatie een automatisch mechanisme is dat de rol van de spreker vergemakkelijkt. Vanuit dit perspectief is accommodatie het product van priming, d.w.z. het proces waarbij blootstelling aan een item de verwerking van dat item in de toekomst vergemakkelijkt. Als een spreker tijdens een gesprek wordt blootgesteld aan een bepaald element van de gesprekspartner, bijvoorbeeld het woord “friet”; zou het voor hen gemakkelijker/efficiënter zijn om dit element opnieuw te gebruiken in de interactie, dan het gebruik van andere vergelijkbare elementen zoals “patat”. Blootstelling aan de spraak van de gesprekspartner zou er daarom toe leiden dat de spreker om efficiëntieredenen dezelfde communicatiepatronen gebruikt als de gesprekspartner.

Accommodatie is ook voorgesteld als een mechanisme dat de rol van de luisterende gesprekspartner in een interactie faciliteert (Clark, 1996). Een spreker kan zijn spraak aan de gesprekspartner aanpassen om ervoor te zorgen dat de gesprekspartner de boodschap begrijpt. Een spreker kan er bijvoorbeeld voor kiezen om het woord friet (in plaats van patat) te gebruiken als hij denkt dat dit woord beter zal worden begrepen door de gesprekspartner. Er zou dus aanpassing plaatsvinden om begrip te faciliteren en miscommunicatie te voorkomen.

Eerdere literatuur lijkt erop te wijzen dat het belang van deze twee perspectieven sterk afhangt van de omstandigheden van het gesprek, bijvoorbeeld of de spreker zijn moedertaal of een tweede taal gebruikt (Costa et al., 2008) en of de gesprekspartner een persoon of een computer is (Branigan et al., 2010). Om het mechanisme achter accommodatie te onderzoeken, bestuderen onderzoekers gesprekken tussen twee gesprekspartners vaak in een laboratoriumsetting.

Op enkele uitzonderingen na (Berry & Ernestus, 2018; Romera & Elordieta, 2013), hebben de meeste studies de neiging om de ecologische validiteit van de interactie op te offeren, d.w.z. de spontaneïteit en natuurlijkheid van het gesprek, om volledige controle uit te kunnen oefenen over de interactie die in het lab plaatsvindt. Zo wordt de stem van de gesprekspartner vaak vooraf opgenomen en via een computer aan de deelnemers gepresenteerd. Als gevolg hiervan communiceren de deelnemers in deze experimenten met een computer, in plaats van met een persoon. Bovendien zijn de taken die vaak worden gebruikt in deze experimenten vaak extreem vereenvoudigde versies van interacties, zoals zogenaamde *shadowing tasks*, waarbij deelnemers geïsoleerde woorden moeten herhalen die door de computer worden afgespeeld (Cohn et al., 2019). Deze omstandigheden belemmeren de generaliseerbaarheid van deze settings naar echte menselijke interacties.

Dit proefschrift heeft tot doel onze kennis over fonetische accommodatie in een tweede taal (T2) uit te breiden door te onderzoeken welke mechanismen ten grondslag liggen aan het leren van de T2-uitspraak in interacties. Het proefschrift bestaat uit een inleiding, vier experimentele hoofdstukken en een algemene discussie. De studies beschreven in de experimentele hoofdstukken onderzoeken fonetische accommodatie bij Nederlandstaligen die Engels leren, die deelnemen aan een interactie in het Engels. Zoals hierboven vermeld, heeft deze groep sprekers vaak moeite met de uitspraak van het contrast tussen de Engelse klinkers in bad en bed (/bæd-bɛd/), aangezien ze de neiging hebben om beide klinkers uit te spreken als een tussenklank die meer lijkt op de klinker /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006).

Hoofdstuk 2 presenteert een nieuw experimenteel paradigma, het Buiksprekerparadigma (*Ventriloquist paradigm* in het Engels), ontwikkeld voor het bestuderen van gesprekken in laboratoriumomgevingen. Met dit paradigma communiceren deelnemers met een handlanger, een persoon die onderdeel is van het experiment. De deelnemer en de handlanger zitten in dezelfde ruimte en zijn uitgerust met koptelefoons en microfoons terwijl ze samen een coöperatief spel spelen. De deelnemers denken dat ze via de koptelefoon naar de stem van de handlanger luisteren, maar in werkelijkheid luisteren ze naar vooraf opgenomen spraakfragmenten. Wanneer de handlanger wil ‘spreken’, verbergt ze gedeeltelijk haar gezicht achter een monitor en drukt op een knop op een verborgen toetsenbord om het vooraf opgenomen audiofragment af te spelen. Deze methodologie combineert de controle over de fonetische input die de deelnemer ontvangt en ecologische validiteit.

Dit hoofdstuk bevat een gedetailleerde beschrijving van de implementatie van dit paradigma in twee verschillende experimenten met twee verschillende handlangers. Bovendien laat dit hoofdstuk zien dat de geloofwaardigheid van het paradigma goed is: ongeveer 80% van de deelnemers gaf aan te geloven dat ze rechtstreeks met de gesprekspartner spraken en merkten de manipulatie niet op. Ten slotte bevestigt dit hoofdstuk ook dat dit paradigma efficiënter is in termen van ecologische validiteit dan andere paradigma’s die gewoonlijk worden gebruikt om accommodatie te bestuderen; in het bijzonder die waar menselijke handlangers worden vervangen door computers.

Hoofdstuk 3 beschrijft een studie die onderzoekt of tweedetaalsprekers hun Engelse uitspraak anders aanpassen in interacties met mensen vergeleken met interacties met computers.

Om dit te doen, werden twee identieke opstellingen gemaakt: een waar niet-moedertaalsprekers interacties hadden met een menselijke gesprekspartner en een waar ze een interactie hadden met een computer. Met behulp van het buiksprekerparadigma en vooraf opgenomen spraak, werd de spraak van de handlanger of computer constant gehouden in alle opstellingen. In het bijzonder hebben we onderzocht of specifieke factoren – zoals *blootstelling* aan de uitspraak van de gesprekspartner van een Engelse klank en *feedback* die een misverstand van de gesprekspartner impliceerde – fonetische accommodatie in deze twee situaties op een andere manier teweegbrachten. Om dit te doen, werd de aan- of afwezigheid van corrigerende feedback van de gesprekspartner geïmplementeerd. Het vooraf opnemen van de spraak van de gesprekspartner zorgt voor een strikte controle over de aan- of afwezigheid van blootstelling aan bepaalde geluiden, zoals de klinker in kwestie, in de spraak van de gesprekspartner tijdens het gesprek. Impliciete corrigerende feedback werd ook geïmplementeerd in het experiment. Uit de feedback bleek dat de manier waarop de deelnemer de Engelse klinker /æ/ uitsprak, voor misverstanden zorgde tijdens het gesprek. In het bijzonder deed de gesprekspartner alsof zij woorden begreep met de klinker /ɛ/, bijvoorbeeld “bed”, wanneer de deelnemer woorden zei inclusief de klinker /æ/, bijvoorbeeld “bad”.

De resultaten van dit hoofdstuk, gebaseerd op de akoestische analyses van de betreffende klinkers, laten zien dat Nederlandstaligen die werden blootgesteld aan de oorspronkelijke uitspraak van de klinker /æ/, meer accommoderen wanneer de blootstelling afkomstig was van de menselijke gesprekspartner dan van de computer. Vergelijkbaar was feedback op de uitspraak van de niet-moedertaalspreker effectiever in het uitlokken van accommodatie als het van een persoon kwam dan van een computer. Verrassend genoeg leidde de combinatie van beide factoren, blootstellingen feedback, niet tot accommodatie bij de sprekers die met de menselijke gesprekspartner spraken. In de groep deelnemers die met de computer communiceerden, veroorzaakte de combinatie van de twee factoren meer accommodatie dan wanneer de twee factoren afzonderlijk werden gepresenteerd.

Eerdere literatuur benadrukt het belang van het gebruik van akoestische analyses in combinatie met perceptuele metingen om rekening te houden met de mate van accommodatie. In Hoofdstuk 4 worden drie verschillende perceptuele metingen gebruikt om de effecten van blootstelling aan de uitspraak van de spreker en corrigerende feedback op de mate van accommodatie in de context van interacties tussen mensen interacties te onderzoeken. De drie perceptuele metingen omvatten beoordelingen van verstaanbaarheid en gelijkenis aan de uitspraak van de vreemde taal. Deze metingen die zijn beoordeeld door moedertaalluisteraars van het Engels en beoordelingen van de verstaanbaarheid zijn ook gegenereerd met behulp van automatische spraakuitlijningssoftware.

De resultaten geven aan dat de verstaanbaarheid toenam voor die tweedetaalsprekers die tijdens de interactie werden blootgesteld aan de uitspraak van de gesprekspartner en feedback kregen op hun eigen uitspraak. De verbetering was nog groter wanneer de factoren werden gecombineerd. Ten tweede laten de resultaten zien dat de Engelse uitspraak van de Nederlandstaligen na het gesprek meer als het Engels klonk, alleen wanneer de sprekers afzonderlijk blootstelling of feedback hadden gekregen, maar niet in combinatie. Zoals eerder genoemd, werd de uitspraak van de deelnemers ook geëvalueerd met behulp van automatische spraakuitlijningssoftware. Dit soort programma’s kan een efficiënt alternatief bieden voor menselijke beoordeling van spraakverstaanbaarheid. De resultaten laten zien dat deze software evaluaties van de verstaanbaarheid genereerde die over het algemeen vergelijkbaar waren met die van moedertaal luisteraars. Er werden echter belangrijke verschillen gevonden tussen de twee verstaanbaarheidsmetingen.

Hoofdstuk 5 bespreekt het effect van het vaardigheidsniveau op fonetische accommodatie door tweedetaalsprekers. Eerdere literatuur biedt tegenstrijdige voorspellingen over de vraag of taalvaardigheid accommodatie verbetert of belemmert. Enerzijds, aangezien het gebruik van een tweede taal al een aanzienlijke cognitieve inspanning vereist, zou kunnen worden verwacht dat meer bekame sprekers meer cognitieve middelen tot hun beschikking hebben dan minder bekwame tweedetaalsprekers. Aan de andere kant, als accommodatie wordt opgevat als een fenomeen dat bedoeld is om een succesvolle interactie te verzekeren, zouden sprekers met een lagere vaardigheid, wiens geaccentueerde uitspraak een groter risico kan vormen voor het succes van de interactie, meer accommodatie kunnen tonen dan sprekers met een hoog taalniveau, die over het algemeen een minder sterk accent hebben.

Dit hoofdstuk biedt een heranalyse van de gegevens gepresenteerd in hoofdstuk 3, rekening houdend met het niveau van de Engelse taalvaardigheid van de Nederlandstaligen. In de setting met twee mensen bleek taalvaardigheid de accommodatie te bevorderen: sprekers met een hoge taalvaardigheid accommodeerden meer dan sprekers met een lage vaardigheid. Dit effect werd alleen gevonden in de “controle”-conditie van het experiment, waar deelnemers geen blootstelling of feedback ontvingen. In de setting waar mensen met een computer communiceerden daarentegen bleek taalvaardigheid een belemmering te zijn voor accommodatie: sprekers met een laag taalniveau accommodeerden meer dan sprekers met een hoog taalniveau. Dit effect werd alleen gevonden in de conditie waarin de sprekers zowel feedback kregen over hun uitspraak als blootstelling aan de uitspraak van de spreker.

De hoofdstukken in dit proefschrift documenteren het bestaan van fonetische accommodatie bij tweedetaalsprekers. Accommodatie blijkt zowel een automatisch fenomeen te zijn dat de rol van de spreker vergemakkelijkt, als een proces dat de taak van de luisteraar faciliteert. Daarom ondersteunt dit proefschrift een definitie van accommodatie die beide perspectieven combineert. Bovendien laten de resultaten van deze onderzoeken zien dat de mate van accommodatie sterk afhangt van het type gesprekspartner, namelijk of het een mens of een computer is. Accommodatie aan computers is ook gedocumenteerd in deze hoofdstukken, maar de belangrijkste verschillen die zijn gevonden tussen accommodatie aan computers en accommodatie aan mensen laten zien dat computergebaseerde methodologieën niet direct generaliseerbaar zijn naar interacties tussen mensen.

Resumen en castellano

En el proceso de aprendizaje de una segunda lengua (L2), la pronunciación suele ser uno de los aspectos más difíciles de dominar. De hecho, el acento de los hablantes no nativos de una lengua a menudo presenta rasgos de su lengua materna (L1). Por ejemplo, muchos hablantes de lengua holandesa tienen problemas con la pronunciación de algunos sonidos de la lengua inglesa. Este grupo de hablantes no nativos suele pronunciar la consonante inglesa /θ/, en palabras como “**th**ink”, como una /d/ o /t/ (Hanulíková & Weber, 2012; Wester et al., 2007), y la vocal /æ/, como en “b**a**d”, como una /ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006). Estas dificultades suelen emerger de las diferencias entre los inventarios de sonido de la L1 y la L2. Los sonidos ingleses de los ejemplos anteriores, /θ/ y /æ/, no existen en neerlandés; por lo tanto, estos hablantes tienden a sustituirlos por otros sonidos que sí existen en su L1: /d/ o /t/, y /ɛ/, respectivamente.

Una forma de superar estas dificultades es mediante el uso de la L2 en las interacciones. Durante una conversación, tendemos a adaptar nuestra forma de hablar al interlocutor, es decir, a la persona con la que interactuamos. Por ejemplo, los hablantes de variedades dialectales estigmatizadas tienden a cambiar su forma de hablar cuando interactúan con un hablante de una variedad diferente (menos estigmatizada). Este fenómeno, conocido como *acomodación* (Giles & Ogay, 2007), puede tener repercusión sobre distintos niveles de la lengua, como la selección léxica (Brennan & Clark, 1996), las estructuras sintácticas (Branigan et al., 2000) o la pronunciación (Pardo, 2006). Debido a la acomodación, las diferencias en las formas de expresarse de los interlocutores acaban difuminándose. La acomodación podría ser un elemento motivador del aprendizaje de lenguas: si un hablante no nativo interactúa con un hablante nativo, el hablante no nativo puede acomodar su pronunciación para sonar de forma más similar al interlocutor y, en consecuencia, aprender a usar una pronunciación más parecida a la nativa después de la interacción. Si estas adaptaciones se conservan después de la conversación, la acomodación podría suponer el aprendizaje de la pronunciación de una L2.

El origen de la acomodación, es decir, por qué se acomoda, ha sido el objeto de interés central de la bibliografía especializada. Normalmente se investiga qué factores sociales y lingüísticos motivan la presencia de acomodación en las interacciones. Por ejemplo, la edad del hablante (Zellou et al., 2021), el papel que tiene en la conversación (Pardo, 2006) y la relación de poder entre los dos interlocutores (Willemyns et al., 1997) pueden influir en el grado y la orientación de la acomodación. Desde una perspectiva social, la acomodación se interpreta como un mecanismo para cerrar las brechas sociales entre los interlocutores (Giles & Ogay, 2007). Desde el punto de vista lingüístico, se ha propuesto que este fenómeno es el producto de la eficiencia (Pickering & Garrod, 2004), ya que puede facilitar el papel del hablante o el papel del interlocutor en la interacción. En esta tesis investigamos qué factores lingüísticos pueden provocar que los hablantes no nativos acomoden su pronunciación durante una interacción.

Pickering y Garrod (2004) propusieron que la acomodación es un mecanismo automático que facilita el papel del hablante. Desde esta perspectiva, la acomodación es el producto del *priming*, es decir, el proceso por el cual la exposición a un elemento facilita el procesamiento de dicho elemento en el futuro. Si, durante una conversación, un hablante recibe exposición a un elemento por parte del interlocutor, por ejemplo, la palabra *aceituna*; le sería más fácil y más eficiente usar esta palabra que otras palabras similares, como *oliva*, cuando vuelva a intervenir en la conversación. Por lo tanto, la exposición al habla del interlocutor llevaría al hablante a acomodar y a utilizar los mismos patrones comunicativos que el interlocutor por razones de eficiencia.

Asimismo, se ha propuesto que la acomodación es un mecanismo que facilita el papel del interlocutor, esto es, de la persona que escucha en una interacción (Clark, 1996). El hablante puede acomodar su discurso al interlocutor para asegurarse de que el interlocutor entienda el mensaje. Por ejemplo, un hablante puede usar la palabra *aceituna* (en lugar de *oliva*) si cree que es más probable que el interlocutor entienda esta palabra. Por lo tanto, el ajuste se realizaría para garantizar la comprensión y evitar malentendidos.

Los estudios previos señalan que la ponderación de estas dos perspectivas depende en gran medida de las circunstancias de la conversación. Entre otros factores, puede resultar decisivo saber si el hablante está usando su lengua materna o una segunda lengua (Costa et al., 2008), o bien si el interlocutor es una persona o un ordenador (Branigan et al., 2010). Para investigar los mecanismos subyacentes a la acomodación, normalmente se recrean conversaciones entre dos interlocutores en un laboratorio. Salvo contadas excepciones (Berry & Ernestus, 2018; Romera & Elordieta, 2013), la mayoría de los estudios tienden a sacrificar la validez ecológica de la interacción, es decir, la espontaneidad y naturalidad de la conversación, para poder ejercer un control exhaustivo sobre la interacción, que tiene lugar en un laboratorio. Por ejemplo, para poder controlar las características acústicas del habla del interlocutor, a menudo se graba la voz del interlocutor con antelación y se presenta a los participantes utilizando un ordenador con altavoces o auriculares. En consecuencia, los participantes de estos experimentos no interactúan directamente con una persona, sino que interactúan con un ordenador. Además, las tareas que se utilizan en estas recreaciones de conversaciones que tienen lugar en un laboratorio tienden a ser tareas extremadamente simples, como *shadowing tasks*, en las que los participantes solo tienen que repetir palabras aisladas reproducidas por un ordenador (Cohn et al., 2019). Estas circunstancias dificultan notablemente la extrapolación de los resultados de experimentos en laboratorios a las interacciones entre personas que tienen lugar en la vida real.

Esta tesis tiene como objetivo ampliar nuestro conocimiento sobre la acomodación fonética en hablantes no nativos, investigando qué mecanismos subyacen al aprendizaje de la pronunciación de L2 en contextos interactivos. La tesis consta de una introducción, cuatro capítulos experimentales y un apartado de conclusiones generales. Los estudios descritos en los capítulos experimentales investigan la acomodación fonética de hablantes holandeses cuando participan en una interacción en inglés. Como se ha mencionado anteriormente, este grupo de hablantes tiende a tener problemas con la pronunciación del contraste entre las vocales inglesas en “bad” y “bed” (/bæd/ y /bɛd/), ya que tienden a pronunciar ambas vocales como un sonido intermedio más similar a la vocal / ɛ/ (Elsendoorn, 1985; Wang & van Heuven, 2006).

El Capítulo 2 presenta un nuevo paradigma experimental, el paradigma del Ventrílocuo, desarrollado para el estudio de conversaciones en laboratorio. En este paradigma, los participantes interactúan con una persona cómplice del experimento. El participante y la cómplice se encuentran en la misma sala, equipados con auriculares y un micrófono, mientras participan juntos en un juego cooperativo. El participante cree que escucha la voz de la cómplice a través de los auriculares, pero, en realidad, escucha fragmentos de audio pregrabados. Cuando la cómplice quiere “hablar”, oculta parcialmente su rostro detrás de un monitor y pulsa una tecla en un teclado, situado fuera de la vista del participante, para reproducir un fragmento de audio pregrabado. Esta metodología combina con éxito el control exhaustivo sobre las características acústicas del habla que el participante escucha durante la conversación y la validez ecológica de la interacción.

Este capítulo incluye una descripción detallada de la implementación de este paradigma en dos experimentos diferentes con dos cómplices diferentes. Además, se demuestra que la implementación del paradigma es convincente: aproximadamente el 80 % de los participantes afirmaron creer que estaban hablando directamente con la cómplice y no notaron la manipulación del experimento. Finalmente, este capítulo también confirma que este paradigma es más eficiente en términos de validez ecológica que otros paradigmas comúnmente utilizados para estudiar la acomodación; en particular, aquellos en los que se sustituye al interlocutor humano por un ordenador.

El Capítulo 3 describe un estudio que investiga si los hablantes no nativos acomodan su pronunciación en inglés de manera diferente cuando interactúan con una persona o con un ordenador. Para ello, se crearon dos escenarios idénticos: uno donde los hablantes no nativos interactuaban con una persona que hacía de cómplice del experimento y otro donde interactuaban con un ordenador. Utilizando el paradigma del Ventrílocuo para implementar audio pre-grabado, el habla del interlocutor, persona u ordenador, era idéntico en ambos escenarios. En particular, investigamos si factores específicos, como la *exposición* a la pronunciación del interlocutor y recibir *feedback* del interlocutor sobre la pronunciación no nativa, desencadenan diferentes grados de acomodación fonética dependiendo de si el interlocutor es una persona o un ordenador. Para ello, se implementó la presencia o la ausencia de exposición a la pronunciación del interlocutor y de *feedback* correctivo por parte del interlocutor. Grabar con antelación el habla del interlocutor nos permite controlar de forma estricta la presencia o la ausencia completa de exposición a determinados sonidos, como la vocal en cuestión (/æ/), en el habla del interlocutor. El *feedback* era implícito y correctivo: el interlocutor fingía entender palabras con la vocal /ɛ/, por ejemplo, "bed", siempre que el participante decía palabras con la vocal /æ/, por ejemplo, "bad" durante la conversación.

Los resultados de este capítulo, basados en los análisis acústicos de las vocales de interés, muestran que los hablantes que recibieron exposición a la pronunciación nativa de la vocal /æ/ acomodaron más cuando la exposición provino del cómplice que del ordenador. De forma similar, el *feedback* sobre la pronunciación del hablante no nativo provocó un mayor grado de acomodación cuando provenía de una persona que de un ordenador. Sorprendentemente, la combinación de ambos factores juntos en la conversación, exposición y *feedback*, no motivó la acomodación en los hablantes que interactuaron con el interlocutor humano. En el grupo de participantes que interactuó con el ordenador, la combinación de los dos factores provocó más acomodación que cuando los dos factores se presentaron de forma aislada.

La bibliografía especializada destaca la importancia de utilizar análisis acústicos en combinación con medidas de percepción para medir el grado de acomodación (Pardo et al., 2013). En el Capítulo 4, se utilizan tres medidas perceptuales diferentes para investigar los efectos de la exposición a la pronunciación del hablante y el *feedback* correctivo sobre el grado de acomodación en el contexto de las interacciones entre personas. Las tres medidas incluyen la inteligibilidad, el grado de semejanza a la pronunciación nativa, ambas medidas por oyentes nativos de inglés, y juicios de inteligibilidad generados por un *software* de alineación automática del habla. Estos programas pueden ofrecer una alternativa más eficiente para medir la inteligibilidad que la metodología basada en recabar juicios de oyentes nativos.

Los resultados indican que la inteligibilidad aumentó para aquellos hablantes no nativos que recibieron exposición a la pronunciación del interlocutor y *feedback* sobre su pronunciación durante la interacción. La mejora fue aún mayor cuando se combinaron los factores. En segundo lugar, los resultados muestran que la pronunciación de los participantes sonaba más nativa, solo si los hablantes habían recibido exposición o *feedback* por separado, pero no en combinación. Además, el *software* de alineación automática de habla generó evaluaciones de inteligibilidad similares a las producidas por oyentes nativos. Sin embargo, se encontraron diferencias clave entre las medidas de inteligibilidad humanas y artificiales.

El Capítulo 5 analiza el efecto del nivel de competencia en una segunda lengua en el grado de acomodación fonética en hablantes no nativos. Los estudios previos ofrecen predicciones contradictorias sobre si el nivel de dominio de un idioma facilita o dificulta la acomodación. Por un lado, dado que el uso de un segundo idioma ya requiere un esfuerzo cognitivo considerable, se podría esperar que los hablantes con un nivel más alto tengan más recursos cognitivos disponibles para acomodar que hablantes menos competentes. Por otro lado, si la acomodación se entiende como un fenómeno diseñado para asegurar la comprensión del interlocutor, los hablantes con menor dominio de la segunda lengua, cuya pronunciación (menos parecida a la de los hablantes nativos) puede suponer un mayor riesgo para el éxito de la interacción, podrían presentar más acomodación que los hablantes con mayor dominio, que generalmente tienen una pronunciación más parecida a la nativa.

Este capítulo ofrece un nuevo análisis de los datos presentados en el Capítulo 3, teniendo en cuenta diferentes medidas de la competencia en la L2. En las interacciones entre personas, el nivel de dominio de la L2 facilitaba la acomodación: los hablantes con nivel alto acomodaban su pronunciación más que los hablantes con nivel más bajo. Este efecto solo se encontró en la condición "control" del experimento, donde los participantes no recibieron ni exposición ni *feedback*. Para los participantes que interactuaron con el ordenador se encontró el efecto contrario: el nivel de dominio en la L2 dificultaba la acomodación. En otras palabras, los hablantes con menos competencia acomodaron más que los hablantes con un nivel más alto de inglés. Este efecto solo se encontró en la condición en la que los participantes recibieron *feedback* sobre su pronunciación y también exposición a la pronunciación nativa.

Los capítulos incluidos en esta tesis doctoral documentan la presencia de acomodación fonética en hablantes no nativos. Además, estos estudios demuestran que la acomodación es tanto un fenómeno automático, que facilita el papel del hablante, como un proceso producto de diseñar el habla a la persona que escucha, que facilita la tarea del oyente. Por lo tanto, esta tesis apoya una definición de acomodación que combine ambas perspectivas. Asimismo, los resultados de estos estudios demuestran que el grado de acomodación depende directamente del tipo de interlocutor, esto es, de si es una persona o un ordenador. Finalmente, estos estudios documentan la existencia de acomodación a ordenadores; notablemente, las diferencias clave encontradas entre la acomodación a ordenadores y la acomodación a personas indican que los resultados de estudios que emplean metodologías basadas en interacciones con ordenadores para estudiar la acomodación no son directamente extrapolables a las interacciones entre personas.

Acknowledgements[1]

I have recently become obsessed with Agatha Christie's novels. In some of her (I believe) best novels, the main character is Hercule Poirot, a Belgian detective who always manages to solve the most complicated of crimes. Although I admire and envy his intellect, order and methods –as portrayed by Christie–, there is one aspect about him that I pity: since he usually finds himself in the middle of a murder case, everyone around him can be considered a suspect and, therefore, he has to work all alone. At some point during my PhD track, I realized this dissertation was my own little "mystery" that needed solving and, unlike Poirot, I could not have finished this work without the help of everyone around me. In this section, I would like to offer these people some lines as an insufficient way to thank them.

Most importantly I want to thank my promotors, **Mirjam Ernestus**, and **Mirjam Broersma**. The *Mirjams* constantly guided me during my PhD journey, offering invaluable advice about research and also about my personal development. **Mirjam Broersma** invented the Ventriloquist paradigm, a project that fascinated me the second I heard about it. She successfully obtained funding from the Dutch Research Council (NWO) and carefully recruited the perfect combination of people to develop this paradigm. This project was her VIDI grant, but she gave the people working on it so much freedom that we also felt the project was our own. Mirjam welcomed me to the Netherlands; during my PhD, she let me find my own path and ensured I stayed on it, and she also introduced me to brainstorming sessions.

Mirjam Ernestus basically transformed me as a researcher. She provided me with help, guidance and all the resources available to develop my research skills, including programming and statistics. Although she is an extremely busy person, she always had time dive deeply into the depths of my project in our meetings, helping me interpret the output of a 5-way(!) interaction or giving me very detailed, pers-

---

[1] I am very sorry I ended up going with the long, emotional version of the acknowledgements and now you have to read through all of this.

pective-changing feedback on my manuscripts. The first time I met Mirjam during my job interview, I was a bit intimidated by her Dutch directness (a concept unknown to me until that interview). Nowadays, Mirjam's directness is one of the qualities I value most about her. I am also extremely grateful for her help and support, especially in the very final months of the dissertation. If writing a dissertation is like running a marathon, Mirjam Ernestus is every stand of free water along the way.

Special thanks to the manuscript committee, **Rob Schoonen**, **Marc Swerts**, **Núria Sebastián Gallés**, **Marije Michel** and **Catia Cucchiarini**, for interesting and thoughtful comments on the work presented in this dissertation.

Doing a PhD can be an isolating experience for some; in my case, I was lucky enough to share every step of the way with **Emily Felker.** Emily worked on the perception part of the same project, while I worked on the production part. Collaborating with her made the process of developing the paradigm fun rather than scary. I have learned a lot from working with her: she is an extremely hard-working and detail-oriented person and an excellent writer. Emily is also a great listener and the perfect officemate to share the everyday experiences and adventures of a PhD.

The successful development of the Ventriloquist paradigm was only possible thanks to the involvement of a great team. **Bob Rosbag**, the technical support of the Centre for Language Studies lab, invested a lot of his time and programming expertise in the implementation of a very complicated set-up. He took the time to teach me everything I know about laboratory recording set-ups and he also taught me to distinguish cables. **Margret van Beuningen**, the manager of the lab, was extremely helpful with the arrangements necessary to test our experiments. Special thanks to **Saraodh Favier** who let me borrow her voice for my version of the Ventriloquist Paradigm. I would also like to acknowledge the crucial support of the thesis students and assistants. **Marjolein van Os** and **Elisabeth Süß** provided useful insights during the very first implementations of the paradigm, and **Lilian Ye** enthusiastically helped develop the newest, even more convincing(!), version of the paradigm. In addition, we also were lucky to receive help from **Anna**, **Ellen, Kyra**, **Lisanne**, **Maaike**, **Nikki**, **Ronny**, **Sjoerd**, **Radiah** and **Tahnee**.

My paranymphs (alphabetically) **Katherine Marcoux** and **Lotte Eijk**, also deserve a lot of credit for their support. Katherine and I started in the world of forced aligners together, which made this adventure more entertaining. She also has the ability of making everyone around her feel welcomed and she is practically half Spanish, so being with her always made me feel like home. She always has nice book recommendations (mostly for very depressing books) and playing board games with her is always exciting. I am happy that the fact that we are both highly competitive people has not posed a problem to a friendship (yet). Lotte and I shared an unconditional love for accommodation (although she would rather call it *alignment*). Discussing my studies with her always gave me clarity and new perspectives. She also happens to be one of the most fun people I have ever met. These past years I have really benefited from her need to propose spontaneous plans and her amazing skills to plan trips. I am extremely grateful that my time as a PhD coincided with them and that we get to share each other's company in coffee breaks, meals and sports. We also lived some adventures together abroad, like the ICPhS experience in Melbourne (thanks for letting me have the bigger bedroom with my own bathroom). Thank you both for making the PhD part of my life easier, and the fun part of my life more enjoyable.

Being able to do a PhD in Nijmegen, specifically at the **Centre for Language Studies**, is actually a very privileged experience because you are surrounded by incredibly interesting people. I would like to thank everyone in the **Speech Comprehension and Production** research group, who did not storm out of the room when I mentioned the 5-way interaction in my presentations but, instead, stayed and gave incredibly insightful feedback. Special thanks to **Esther Janse**, who always comes up with interesting questions even though she has listened to my presentations multiple times. I would also like to thank

Louis ten Bosch, for his invaluable help and discussions about different types of regression models and force aligners; and **Roeland van Hout**, for his input on vowels. I was also lucky to be part of the **International Max Planck Research School for Language Sciences** network. Special thanks to **Kevin Lam** for arranging all of the courses, lectures and social activities and for being so approachable. I would also like to thank the PhD candidates that made my journey more pleasant, including **Annika**, **Aurélia**, **Chantal**, **Chen**, **Elly**, **Ferdy**, **Figen**, **Gert-Jan**, **Hanno**, **Lisa**, **Maria**, **Mónica**, **Patricia**, **Saskia**, **Tashi**, **Theresa**, **Tim**, **Xiaoru**, **Wei** and **Yu**, among others. Special thanks to **Claire Goirot**, who I met at the very beginning of my journey as a PhD and who has always listened empathetically to my stories, interpreted for me at the hairdresser's and, most importantly, has been (and is) a mentor to me. I would also like to thank the gezelligheid that reigned in the office I shared with **Emily Felker**, **Lidy Zijlmans**, **Martijn Bentum** and **Toni Rietveld**.

In 2019, I had the wonderful opportunity to do a research stay at Kansas University in Lawrence (Kansas, USA), hosted by **Annie Tremblay**. This visit was incredibly productive: I tested 136 participants for two perception experiments in 18 days thanks to Annie. The perception experiments are one of the most informative parts of my dissertation, so I am really grateful that Annie made it possible: she carefully planned every step of my visit, walking me through the confusing paperwork required to test experiments in the USA, and recruiting and scheduling all my participants. During this trip, I also got to meet the vivid linguistic scene in Lawrence and met people like **Joan Sereno**, **Allard Jongman**, **Caitlin Couhglin**, **Charlie Redmon**, **Seulgi Shin**, among many others. Special thanks to **Nick Feroce** who made sure I felt welcomed and took me to all the right places to visit in Kansas.

During the last years of my PhD, I was given the opportunity to teach at the **Department of Language and Communication**, for which I am extremely grateful. I would like to thank the **Spanish team** that I work with for supporting me in the last stages of my PhD, specially **Núria Domínguez**, **Lieke van Maastricht** and **Andreu van Hooft**. Gracias en especial para **Núria** y **Lieke**, que desde el principio me ayudaron a familiarizarme con el sistema y sacrificaron mucho tiempo para que pudiera centrarme en acabar la tesis. Núria, además, siempre tiene la respuesta para todas las preguntas que tengo y también para las preguntas que todavía no sé que tengo. Gracias también a Lieke, que durante esta última etapa de la maratón siempre me ha animado a celebrar las pequeñas victorias. I would also like to express my gratitude to my new colleagues at the **Faculty of Social Sciences**, the **DOC team**, including **Eliane Segers**, **Marco van de Ven**, **Jos Keuning**, **Liesbeth Crajé-Tilanus**, **Naomi van Berg** and **Anouk Bakker**. Seeing a new project in the horizon definitely gave me the last boost I needed to finish this marathon.

I believe my life has considerably improved ever since I arrived in the Netherlands, and for that I am extremely grateful to **María Luisa Lecumberri**. She told me about this PhD position and encouraged me to apply for it. Thanks also to **Gorka Elordieta**, my MA thesis supervisor, for introducing me to the field of accommodation. And **Myriam Uribe-Etxebarria** for giving me the opportunity to develop as a researcher within the HITT group.

Early in my PhD I realized that the best thing I could do –after 8 hours of being a ventriloquist, (technically) "deceiving" people into thinking I was a computer or analyzing vowels– was sports. Being a part of sports team made me have contact with the real world and helped me understand that being a PhD candidate was only a part of who I am. Ik wil mijn voetbalteam, **Orion VR3**, bedanken voor hun gastvrijheid en voor het helpen bij het verbeteren van mijn voetbalvaardigheden en ook mijn Nederlands. I would also like to thank my **padel teammates** for getting me out of my house during corona times.

I would like to offer special thanks to **Blanca Arias Badia** and **Ander Egurtzegi**. Being a first generation academic, I am grateful I can count on Blanca and Ander as my "older academic siblings", who always have wise words of advice when I face difficult decisions. **Hannah Lutzenberger** and **Rehana Omardeen** also deserve credit for their support. Rehana introduced me to the world of football and true crime podcasts, and she is the best liming companion. Hannah has become one of the most caring friends I have even though, at first, I also did not know whether she liked me (see Lutzenberger, 2022). I am extremely lucky to be part of their adventures and to be invited to all of their quizzes. I would also like to thank my very supportive friends such as **Sergio López Sancio**, **Inés**, **Nils**; **Laura**, **Luis, Natalia**, **Rubén**, and the rest of my **former colleagues from the Basque Country**; **Sergio López Martínez** and **Celia**, who both came to spend some time in Nijmegen; and **Chris**, **Esther**, **Luuk**, **Noralie** and **Maaike** for the delicious dinners.

I would like to thank **the Richters**, omdat ze mijn ondersteunende Nederlandse familie zijn. Por supuesto, gracias a **mi familia** y, en concreto, a mi piña: **Aurora** (mi madre), **Patri** y **Mauri**. Me siento afortunada de contar con vuestro apoyo incondicional. A mi madre le agradezco todo el sacrificio para que pudiéramos salir adelante. Patri es el origen de mi interés por la acomodación fonética. Mauri le ha devuelto la vida a este texto, probablemente convirtiéndose en la única persona que ha leído todas las páginas de este libro.

Lastly, I would like to thank **Anke**, my partner, and **Rosso**, our cat, who have improved my life significantly ($p = <0.001$, although they both dislike p-values). I am extremely lucky to have Anke because she always supports me in every aspect of my life, keeps me alive feeding me delicious food, listens to me ranting about vowels (even though she is convinced that “vwls dn’t mttr!”) and she is the most intelligent, hilarious, understanding and caring person I know. Thank you.

Curriculum Vitae

Aurora Troncoso Ruiz was born in 1991 in Jerez de la Frontera (Cádiz, Spain). In 2014, she obtained a Bachelor's degree in English Philology from the University of Oviedo (Asturias, Spain) and in 2015, she obtained a Master's degree on Linguistics with a specialization in Phonetics and Syntax at the University of the Basque Country (Spain). During her masters, she was employed as a research assistant in the INSPIRE project, working with María Luisa Lecumberri. After her masters, she was employed as a researcher in the Hizkuntzalaritza Teorikorako Taldea, working with Gorka Elordieta in the field of prosodic accommodation. In 2016, Aurora started her PhD within the Speech Production and Comprehension group at Centre for Language Studies (Radboud University) The results of this research project are described in this dissertation. During this project, she was a visiting scholar at the University of Kansas (USA) in 2019, hosted by Annie Tremblay. In addition, she has been involved in teaching positions as a lecturer at Radboud University. Currently, she is a postdoctoral researcher at the Behavioural Science Institute at the Faculty of Social Sciences and a lecturer at the Language and Communication department at the Faculty of Arts (Radboud University).

List of Publications

Felker, E., Troncoso-Ruiz, A., Ernestus, M., & Broersma, M. (2018). The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America*, 144(4), EL304– EL309.
https://doi.org/10.1121/1.5063809

Troncoso-Ruiz, A., & Elordieta, G. (2018). Prosodic accommodation and salience: The nuclear contours of Andalusian Spanish speakers in Asturias. *Loquens*, 4(2), 043.
https://doi. org/10.3989/loquens.2017.043

Troncoso-Ruiz, A., Ernestus, M. & Broersma, M. (2019). Learning to produce difficult L2 vowels: the effects of awareness-rasing, exposure and feedback. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, 1094-1098. doi:10.5007/2175-8026.2018v71n3p99.

To be submitted

Troncoso-Ruiz, A., Ernestus, M. & M. Broersma (to be submitted). Comparing phonetic accommodation in interactions with humans and with computers: effects of exposure and feedback.

Troncoso-Ruiz, A., Ernestus, M. & M. Broersma (to be submitted). Perceiving phonetic accommodation: the effect of exposure and feedback on intelligibility and nativelikeness.