Learning second language speech perception in natural settings

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

ISBN: 978-94-6093-375-2
DOI: https://dx.medra.org/10.48273/LOT0590
NUR: 616

Learning second language speech perception in natural settings

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 van decanen
in het openbaar te verdedigen op dinsdag 10 juni 2021
om 16.30 uur precies

door
Emily Rose Felker
geboren op 9 januari 1991
te Peoria (Verenigde Staten)

Promotor: Prof. dr. M.T.C. Ernestus

Copromotor: Dr. M.E. Broersma

Manuscriptcommissie:

Prof. dr. M. van Oostendorp
Prof. dr. M.L.G. Lecumberri (Universidad del País Vasco, Spanje)
Prof. dr. W.M. Lowie (Rijksuniversiteit Groningen)
Dr. N.H. de Jong (Universiteit Leiden)
Dr. E. Reinisch (Österreichische Akademie der Wissenschaften, Oostenrijk)

This research was funded by a Vidi grant (276-89-006) awarded
by the Dutch Research Council (NWO) to Mirjam Broersma.

Table of Contents

Chapter 1: Introduction

Learning how to listen in a second language is a crucial component of second language acquisition and is vital for successful communication. While speech perception in a native language (L1) is usually an effortless process, second language (L2) speech perception poses many difficulties for the non-native listener, the most notable of which is word recognition (Cutler, 2012). One of the challenges is hearing the difference between two speech sounds that are contrastive in the L2 but not in the L1 (Best, 1994; Best & Tyler, 2007). For instance, native Dutch listeners may confuse the English words "pan" (/pæn/) and "pen" (/pɛn/) since they perceptually assimilate both /æ/ and /ɛ/ into the same phonemic category in their L1. Another challenge for L2 listeners is adapting to regional or foreign accents. For example, a Dutch traveler in Australia may confuse the words "pen" (/pɛn/) and "pin" (/pɪn/) because they are not familiar with the /ɛ/-to-/ɪ/ vowel shift in certain Australian English dialects. An important question in second language acquisition is how L2 listeners can overcome these word recognition challenges to improve their perception of L2 speech.

Traditionally, L2 speech processing and perceptual learning have been studied using non-interactive, computer-based training programs that provide intensive exposure to de-contextualized, highly controlled stimuli (e.g., Bradlow et al., 1999; Iverson & Evans, 2009; Lively et al., 1994; Logan et al., 1991; Sakai & Moorman, 2018). These experimental paradigms have yielded valuable insights about how various aspects of the phonetic input can contribute to perceptual learning. However, they do not address the more natural learning situations that L2 listeners encounter in everyday life, such as learning in conversational interaction. Reconciling control over phonetic input with ecological validity is an important methodological goal for L2 speech learning research: this would allow us to test whether certain L2 speech learning mechanisms, proposed on the basis of research using artificial listening tasks, transfer to real-life communicative settings.

This dissertation aims to make methodological and theoretical contributions to the study of improving L2 speech perception in natural contexts. As an introduction, Section 1.1 will first motivate the need for improving the ecological validity of L2 speech perception research, focusing on the importance of more naturalistic speech stimuli and learning contexts. Section 1.2 will introduce three perceptual learning mechanisms, ranging from relatively implicit to relatively explicit, that are investigated throughout the dissertation: lexical guidance, interactional corrective feedback, and phonetic instruction. Finally, Section 1.3 will present a chapter-by-chapter outline specifying the research questions and methodologies addressed in this dissertation.

1.1 Expanding the scope of research to more natural language processing

In recent years, researchers in psycholinguistics and related fields have called for studying language processing of more natural types of speech in more ecologically valid contexts (e.g., Tanenhaus & Brown-Schmidt, 2008; Tucker & Ernestus, 2016; Willems, 2017). One way to achieve this is to study the processing of continuous speech, rather than isolated words, and to use stimuli from more casual speech registers, such as conversational speech. Conversational speech differs from formal speech not only in syntactic structure and lexical content but also in numerous phonological and phonetic properties (e.g., Tucker & Ernestus, 2016). In particular, reduced pronunciation variants, in which phonemes are weakly articulated or altogether missing, are a hallmark of casual speech (e.g., Ernestus & Warner, 2011; Johnson, 2004) and have been shown to influence the processes involved in speech perception (e.g., Brouwer et al., 2012; Janse & Ernestus, 2011; Kemps et al., 2004). These reduced pronunciation variants often cause speech comprehension problems for L2 listeners (e.g., Brand & Ernestus, 2018; Ernestus et al., 2017). Experiments that employ recordings of continuous speech extracted from real conversations would represent a step forward in the direction of studying the type of speech that L2 listeners are likely to encounter in everyday life.

To investigate the perception of continuous speech, dictation tasks that require listeners to transcribe the words they hear have long been used in speech intelligibility research (Kent, 1992) and in L2 teaching (Field, 2003; Morris, 1983; Oller & Streiff, 1975; Siegel & Siegel, 2015; Stansfield, 1985). However, dictation tasks are typically only scored for accuracy at the word level (Buck, 2001; Kent, 1992; Savignon, 1982), thereby excluding information about what phonemes in particular posed problems for the listener. Only recently have more fine-grained scoring measures, such as those based on phonetic feature similarity between a listener's transcription and a target phrase, come into use in phonetics research (e.g., Podlubny et al., 2018). Chapter 2 demonstrates how the dictation task can become a more valuable tool for research in L2 speech perception by describing and evaluating a range of scoring measures that can be combined to provide more detailed information about what listeners can recover from the speech input and the communicative consequences of their perceptual errors.

Beyond merely studying how L2 listeners perceive casual speech that was produced in a conversational context, a more ambitious goal is to investigate how L2 listeners process speech when they themselves are actively participating in a conversation. Unlike language processing in isolation, language processing in dialogue is influenced by interlocutors' conversational goals and by social-pragmatic cues, including what interlocutors know about each other's perspective, knowledge, and intentions (Tanenhaus & Brown-Schmidt, 2008). Cognitive scientists have proposed that humans are actually "designed" for dialogue, the most natural and basic setting for language use (Garrod & Pickering, 2004), and that largely unconscious interactive alignment mechanisms facilitate language processing and may contribute to both L1 and L2 acquisition (Pickering & Garrod, 2004; Costa et al., 2008). In addition to promoting unconscious linguistic alignment between speakers, dialogue can also bring aspects of language itself to conscious awareness. Conversational interaction, along with the negotiation for shared meaning that it entails, is theorized to be an important locus of language learning within the field of L2 acquisition (Ellis, 1999, 2003; Long, 1980). This is because conversation provides L2 learners with the opportunity for interactional feedback that draws their attention to discrepancies between the actual target language forms and their current knowledge of them (Schmidt, 2001). To date, little is known about the perceptual learning of L2 speech in conversational settings, as the existing interactional L2 acquisition research typically focuses on production rather than perception and addresses the learning of grammar and vocabulary rather than phonology.

The relative lack of research on L2 perceptual learning in conversational interaction likely arises from methodological challenges in experiment design. Traditional research paradigms for studying language in dialogue, such as the Map task (Brown et al., 1983) and Diapix task (Van Engen et al., 2010), allow for semi-spontaneous interaction between interlocutors but, as a result, cannot fully control the language input participants receive. The syntactic and lexical input can be controlled with a confederate scripting paradigm (Branigan et al., 2000), in which a confederate of the experimenter interacts with a participant while saying only pre-determined sentences. However, no matter how well-trained the scripted confederate is, the phonetics of their live speech can never be fully controlled due to the inherent variability in the realization of speech sounds. Pre-recording the confederate's speech would solve this problem. Chapter 3 describes and validates an innovative method called "the ventriloquist paradigm" that incorporates pre-recorded speech in a face-to-face interaction while simultaneously maintaining the illusion of a live conversation, thereby reconciling phonetic control with ecological validity. In Chapter 4, the ventriloquist paradigm is then employed to investigate L2 perceptual learning mechanisms in a dialogue setting.

1.2 Learning mechanisms for second language speech perception

With a focus on natural contexts for L2 speech learning, as motivated above, this dissertation investigates three learning mechanisms for L2 speech perception that span the spectrum from more implicit to more explicit: lexical guidance (Section 1.2.1), interactional corrective feedback (Section 1.2.2), and phonetic instruction (Section 1.2.3).

1.2.1 Implicit lexical guidance

In L1 speech perception, lexically guided perceptual learning is one of the most well studied implicit learning mechanisms for adapting to the enormous variability in the speech signal across speakers and dialects (e.g., McQueen, Cutler & Norris, 2006; Norris et al., 2003). Essentially, listeners use their top-down lexical knowledge to constrain their interpretation of ambiguous speech sounds, leading to long-term adjustments in their phonemic category boundaries. For instance, suppose a speaker's accent involves raised front vowels such that her /ɛ/ pronunciations sound more like /ɪ/. If a listener repeatedly hears that speaker pronounce /ɪ/-like vowels in lexical contexts where /ɛ/ was expected (e.g., pronouncing "went" as /wɪnt/ and "west" as /wɪst/), the listener's phonemic boundary for /ɛ/ adjusts to allow for the /ɪ/-like realizations. Lexically-guided perceptual learning has most often been studied with consonants but has also been demonstrated for ambiguous vowels (McQueen & Mitterer, 2005) and for cross-category vowel shifts (Maye et al., 2008), which distinguish many varieties of English around the world (Thomas, 2001). Moreover, while lexical guidance has primarily been studied in L1 listening (see reviews of Samuel & Kraljic, 2009 and Baese-Berk, 2018), a few studies have shown that it also drives perceptual retuning in L2 listeners (e.g., Cooper & Bradlow, 2018; Drozdova et al., 2016). The fact that foreign-language subtitles improve L2 listeners' ability to perceive speech in unfamiliar regional accents (Mitterer & McQueen, 2009) supports the theory that lexical constraints on the interpretation of phonetic input help to retune speech perception.

The effectiveness of lexical guidance for perceptual learning of speech in conversational interaction, as opposed to in purely receptive listening contexts, has yet to be demonstrated empirically. Since lexically guided perceptual learning is assumed to be an automatic process (McQueen et al., 2006) and is robust to task-based changes in listening strategy (Drouin & Theodore, 2018), one might expect this learning mechanism to work just as well when listeners are also engaged in dialogue. Lexically guided perceptual learning has been shown to occur under more challenging listening conditions, such as when listening with an added memory load or with divided attention (Zhang & Samuel, 2014). On the other hand, several studies using various types of perceptual and lexical training paradigms have shown that when people have to alternately speak and listen during training, their perceptual improvement is impaired (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007). Baese-Berk (2019) speculated that this impaired perceptual learning may result from an overload in shared cognitive processing resources between the speech perception and production modalities. Chapters 4 and 5 investigate lexically guided perceptual learning in the context of a task-based dialogue that is relatively cognitively demanding on multiple fronts: participants have to alternate between solving visual puzzles and both speaking and listening in their L2. The research in these chapters thus goes beyond the research on lexical guidance in the passive listening contexts in which it is traditionally studied, extending it to active communicative contexts representative of real-world cognitive processing demands.

1.2.2 Interactional corrective feedback

The second learning mechanism studied in this dissertation, interactional corrective feedback, is one that has been extensively studied in L2 instructional settings for grammar, vocabulary, and pronunciation learning, though rarely for perceptual learning (e.g., see the meta-analysis of Brown, 2016). Interactional feedback in the language classroom, which often takes the form of clarification requests or recasts, has been shown to promote noticing of L2 forms, which in turn positively affects subsequent learning of those forms (e.g., Mackey, 2006).

Research has shown that recasts, which represent a more implicit type of feedback, are less effective for L2 learning than corrective feedback that is more explicit in nature (Rassaei, 2013). What makes the difference seems to be L2 learners' interpretation of the feedback, including whether they thought the feedback was intended to be corrective and whether they noticed the linguistic target of the feedback (Mackey et al., 2000; Rassaei, 2013). By examining learners' immediate response to the feedback, or uptake (Lyster & Ranta, 1997), researchers can better understand whether interactional feedback was interpreted accurately (Mackey et al., 2000).

Though corrective feedback is typically the mechanism employed in computer-based training programs for the perceptual learning of L2 sound contrasts (e.g., Bradlow et al., 1999; Iverson et al., 2005; Lee & Lyster, 2016b; Wang & Munrow, 2004), corrective feedback for speech perception has only rarely been studied in interactive contexts. One study using simulated classrooms with form-focused instruction did find that perception of novel L2 sound contrasts improved more for learners whose instructor provided corrective feedback in response to their perceptual errors, compared to learners in a control classroom who completed the same sound learning activities without feedback (Lee & Lyster, 2016a). As the corrective feedback in these speech perception studies came from computer programs or a language teacher, the linguistic target and corrective nature of the feedback was very explicit and thus relatively easy to learn from. It remains an open question how interpretable corrective feedback about speech perception would be in a more natural dialogue between an L2 listener and an L1 speaker. Moreover, the effectiveness of interactional corrective feedback for learning to perceive a wholly unfamiliar accent in the L2, as opposed to a novel L2 sound contrast, has never before been tested. Using a task-based dialogue, Chapter 4 investigates two types of corrective feedback, one more implicit and one more explicit, and assesses their effect on L2 listeners' uptake and online processing of the L1 speaker's unfamiliar accent.

1.2.3 Explicit phonetic instruction

The third and most explicit learning mechanism addressed in this dissertation is phonetic instruction, which can naturally be integrated into the L2 classroom and thus may have more real-world relevance than the intensive computer-based training programs used in the speech perception literature (Kissling, 2014). Phonetic instruction can improve listeners' perception of difficult contrasts in a non-native language by drawing their attention to the way words sound, rather than to their meaning (Guion & Pederson, 2007). Even more specifically, instruction that draws non-native listeners' attention to specific sound classes within words, such as vowels versus consonants (Pederson & Guion-Anderson, 2010) or tones versus consonants (Chen & Pederson, 2017), has been shown to specifically improve listeners' perception of the sound classes to which they attended. According to attention-to-dimension models of perceptual learning, the acquisition of unfamiliar phonetic contrasts involves listeners shifting their selective attention to the acoustic-phonetic cues that are relevant for those contrasts (Francis et al., 200; Francis & Nusbaum, 2002). Supporting this theory, multiple studies have shown that explicitly instructing listeners about what specific phonetic cues to attend to, such as phoneme duration (Hisagi & Strange, 2011; Porretta & Tucker, 2014) and tonal pitch height (Chandrasekaran et al., 2016), improves their perception of non-native contrasts.

The existing perception research about using explicit instruction to draw attention to specific phonemes and phonetic cues (Chandrasekaran et al., 2016; Chen & Pederson, 2017; Guion & Pederson, 2007; Hisagi & Strange, 2011; Pederson & Guion-Anderson; Porretta & Tucker, 2014) has focused on instructing non-native listeners about the sounds of an unfamiliar language. While this allows listeners' pre-existing knowledge of the language in question to be controlled, such a learning scenario would likely never occur outside the laboratory. The effectiveness of explicit instruction for advanced L2 listeners, whose phonemic categories may be more entrenched after years of language exposure and usage, remains an open question. Moreover, if phonetic instruction works primarily by re-orienting listeners' attention to the right cues, such instruction may be less effective for elderly adult listeners, who have been shown to have less perceptual flexibility (Scharenborg & Janse, 2013), selective attention capacity (Sommers, 1997), and ability to generalize perceptual learning (see review of Bieber & Gordon-Salant, in press), compared to the younger adults who are traditionally studied in L2 research. Finally, while theories of L2 acquisition posit that conscious awareness and attention to form play a fundamental role (e.g., Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994), the link between phonological awareness of specific L2 sound contrasts and the ability to distinguish those contrasts in perception has not yet been empirically established. To address these gaps in the literature, Chapter 6 investigates the effect of explicit phonetic instruction on both phonological awareness and perception of L2 contrasts for relatively high-proficiency young and older adult L2 listeners.

1.3 Outline and research questions

This body of this dissertation consists of two methodological chapters and three theoretically driven experimental chapters that all serve the aim of studying L2 speech perception in natural learning contexts using well controlled and ecologically valid methods. This section outlines the chapters to come and presents their specific research questions and methodologies.

To show how L2 speech perception can be studied with the use of more naturalistic speech stimuli, such as continuous conversational speech, **Chapter 2** focuses on the dictation task, a well-known tool in second language teaching in which learners have to transcribe short stretches of speech and their transcriptions are scored for accuracy. Excerpts from a casual American English conversation are used to create a dictation task that is administered to American English (L1) and Dutch (L2) listeners. Four different measures that can be used to score the dictation task, each with their own advantages, are presented and compared: lexical error rate, orthographic edit distance, phonological edit distance, and semantic error rate. The chapter's goal is to assess the validity and utility of these measures by analyzing how well they distinguish L1 and L2 listeners, how listeners' performance differs across the measures, to what extent the measures correlate with measures of L2 proficiency and usage, and to what extent the measures correlate with each other.

Going beyond the study of naturalistic stimuli to the study of naturalistic interaction, **Chapter 3** presents the novel "ventriloquist paradigm," an experimental method for studying speech processing in dialogue with full control over phonetic exposure. In this paradigm, a participant interacts face-to-face with a confederate who, unbeknownst to the participant, communicates by using a hidden keyboard to play pre-recorded utterances to the participant's headphones while briefly ducking her face behind a computer screen. The pre-recorded speech, which is designed to meet whatever phonetic constraints are required for the experiment, includes both task-relevant phrases and flexible phrases that can be used to respond to any spontaneous questions from the participant. This chapter aims to describe the ventriloquist paradigm in detail and establish the paradigm's validity by answering the following questions: First, does the ventriloquist paradigm reliably convince participants they are having a genuine conversation? Second, how important is the face-to-face context for making the illusion convincing? Finally, does the ventriloquist paradigm create more engaging and interactive conversation than a comparable setup in which participants believe their interlocutor is a computer?

In **Chapter 4**, the ventriloquist paradigm and the computer-interlocutor control setup are both used to investigate the role of corrective feedback and lexical guidance in the perceptual learning of a novel L2 accent in dialogue. Dutch participants play an information-gap game with an English-speaking interlocutor whose accent exhibits an unexpected vowel shift in which /ɛ/ is pronounced as /ɪ/. Participants can learn about the vowel shift either from implicit lexical guidance built into the game, which constrains their possible interpretation of the interlocutor's words, or from interactional corrective feedback, whereby the interlocutor interjects whenever the participant makes an error indicating that they misperceived the interlocutor's accented pronunciation. Two types of corrective feedback are compared: generic feedback, in which the interlocutor simply points out that an error was made, and contrastive feedback, in which the interlocutor explicitly contrasts the misperceived word with the intended word.

This chapter addresses the following research questions: First, does corrective feedback about erroneous perception of a novel accent in dialogue lead to uptake during the interaction for L2 listeners, and if so, which type of feedback leads to more uptake: generic or contrastive corrective feedback? Second, do corrective feedback and lexical guidance about a novel accent in dialogue improve L2 listeners' online processing of the accent? Third, does the amount of perceptual learning differ between the face-to-face ventriloquist paradigm and the computer-interlocutor control setup? For the first question, uptake is operationalized as word identification accuracy for critical accented words (e.g., "pen" pronounced as /pɪn/) during the course of the game's critical trials. For the second question, online processing of the accent is assessed with a lexical decision task that is presented as a "Word or not?" game with the same interlocutor immediately following the dialogue. In this task, the key question is whether listeners will become faster and more accurate at judging critical words (e.g., "best" pronounced as "bɪst") as being real words. For both questions, participants in the lexical guidance and corrective feedback conditions are compared to control participants who took part in the same task-based dialogue without receiving any evidence for the vowel shift.

Following up on the preceding chapter's results, **Chapter 5** takes a closer look at lexically guided perceptual learning in an interactive L2 dialogue setting. Dutch listeners take part in the same interactive experiment as in

Chapter 4, in either a lexical guidance condition or a control condition, but their perceptual learning is tested differently than before. Instead of using a lexical decision post-test, the experiment investigates perceptual learning with a phonetic categorization pre-test and post-test using a twelve-step vowel continuum between /ɛ/ and /ɪ/. This task evaluates how listeners' phonemic boundaries shift as a result of their experience in the interaction, specifically, whether their /ɛ/ category boundary expands to include more /ɪ/-like realizations in line with interlocutor's accent. This approach sheds more light on how perception changes over time within individual listeners, complementing the preceding chapter's between-groups analyses. As the phonetic categorization task would likely be unconvincing with a face-to-face interlocutor, this study employs only the computer-interlocutor setup.

Next, **Chapter 6** investigates the effectiveness of explicit phonetic instruction for improving phonological awareness and perception of L2 sound contrasts in younger and older adults. Dutch listeners receive a short video instruction about one of two English phonemic contrasts that, due to differences between Dutch and English phonology (e.g., Collins & Mees, 1996), should pose differing degrees of perceptual difficulty: the word-final /t/-/d/ contrast (a familiar contrast in an unfamiliar position, expected to be easier) and the /æ/-/ɛ/ contrast (a completely unfamiliar contrast, expected to be more difficult). For each contrast, the video instruction either does or does not describe how the phonetic cue of vowel duration can be used to distinguish the contrast. Listeners' phonological awareness and perception of each contrast are assessed at pre-test and post-test. Awareness is operationalized as the degree of knowledge that members of minimal word pairs based on a contrast are meant to sound different, and this is measured with a task in which participants make "same" or "different" judgments to a series of visually presented word pairs that are either minimal pairs or homophones. Perception is measured in a two-alternative forced-choice listening test with the critical /t/-/d/ words and /æ/-/ɛ/ words from the awareness task. The main research questions are as follows: What is the relationship between phonological awareness and perception of novel L2 sound contrasts? Can explicit phonetic instruction increase phonological awareness and perception of novel L2 sound contrasts for young and elderly listeners? If so, does learning increase if the instruction describes a phonetic cue that distinguishes the contrasts? The analyses also assess the extent to which any learning effects observed differ between the two sound contrasts and between the two age groups.

Finally, **Chapter 7** summarizes and provides a general discussion of the results of the studies in the preceding chapters. The discussion makes methodological recommendations for studying L2 speech perception in more natural contexts; synthesizes the results from the studies about lexical guidance, interactional corrective feedback, and phonetic instruction; discusses practical implications of the present findings; and lays out questions for future research.

Chapter 2: Evaluating dictation task measures for the study of speech perception

Abstract:
This paper shows that the dictation task, a well-known testing instrument in language education, has untapped potential as a research tool for studying speech perception. We describe how transcriptions can be scored on measures of lexical, orthographic, phonological, and semantic similarity to target phrases to provide comprehensive information about accuracy at different processing levels. The former three measures are automatically extractable, increasing objectivity, and the middle two are gradient, providing finer-grained information than traditionally used. We evaluate the measures in an English dictation task featuring phonetically reduced continuous speech. Whereas the lexical and orthographic measures emphasize listeners' word identification difficulties, the phonological measure demonstrates that listeners can often still recover phonological features, and the semantic measure captures their ability to get the gist of the utterances. Correlational analyses and a discussion of practical and theoretical considerations show that combining multiple measures improves the dictation task's utility as a research tool.

This chapter is based on the following:
Felker, E., Ernestus, M. & Broersma, B. (2019). Evaluating dictation task measures for the study of speech perception. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th ICPhS*, Melbourne, Australia 2019 (pp. 383–387). Canberra, Australia: Australasian Speech Science and Technology Association In

2.1 Introduction

One of the most straightforward ways to test how accurately listeners can decode the acoustic speech signal into linguistic units, such as words, is to have them transcribe a stretch of speech. In the field of applied linguistics, this method is known as the dictation task, and we argue in this paper that the dictation task has untapped potential as a phonetics research tool for the study of speech perception.

In second language (L2) learning and teaching, the dictation task is widely used both as a pedagogical tool and as a testing instrument for listening skills (Matthews & O'Toole, 2015; Morris, 1983; Oller & Streiff, 1975; Stansfield, 1985). The dictation task is particularly relevant for training and evaluating perceptual processing abilities, such as phoneme recognition and lexical segmentation (Field, 2003; Siegel & Siegel, 2015). Despite the ubiquity of the dictation task in language education, however, it has seen relatively little use in the field of phonetics, even though written transcriptions of speech are often used in the context of speech intelligibility research (Kent, 1992).

An important reason why dictation is underutilized in phonetics research may be that detailed scoring measures have yet to be developed. In applied linguistics, transcriptions in dictation tasks are usually scored for word- or phrase-level accuracy, with potential latitude given by human raters for misspellings (Buck, 2001; Savignon, 1982). The percent of words correctly identified is also a typical scoring measure in the field of speech intelligibility testing (Kent, 1992). However, examining only the proportion of words accurately transcribed does not differentiate completely wrong and more nearly right answers. Consider the utterance "my Friday night" spoken with the consonants not clearly articulated, which one listener transcribes as "my friend and I" and another as "my family" in the experiment we report. Both answers match the target phrase in exactly one word, but the former is a better phonological match. Binary measures like word error rate ignore finer distinctions between answers at the phonological level, such as how well listeners can recover the target words' phonetic features.

We propose that considerable information about listeners' perceptual abilities can be gained by scoring transcriptions with a broader range of measures that capture accuracy at different processing levels. Moreover, using automatically calculated measures increases scoring objectivity. Finally, complementing word-, letter-, and phoneme-based measures with a semantic accuracy measure provides insight into the communicative consequences of perceptual errors.

This paper demonstrates how a dictation task with more precise measures can be used to study speech perception. Specifically, we present four measures—lexical error rate, orthographic edit distance, phonological edit distance, and semantic error rate—and evaluate their usefulness when applied to a dictation study investigating how non-native listeners perceive casual speech with severe speech reductions.

Speech reductions, in which segments and even syllables are weakly articulated or altogether missing, are a hallmark of the casual speech register (Ernestus & Warner, 2011; Johnson, 2004). While native (L1) listeners can easily process reduced words presented in context (e.g., Ernestus et al., 2002; Janse & Ernestus, 2011; Kemps et al., 2004), reductions often cause comprehension problems for non-native listeners, who tend to have less exposure to these pronunciation variants (Brand & Ernestus, 2018; Ernestus et al., 2017).

We tested Dutch non-native and American English native listeners on a fill-in-the-blank dictation task with American English target phrases containing massive phonetic reductions, presented in sentential contexts. To evaluate the four dictation measures, we analyze how well they distinguish the listener groups, how performance differs across the measures, how the measures correlate with the non-natives' language proficiency and usage, and how the measures correlate with each other. Following these analyses, we discuss the measures' utility based on practical and theoretical considerations.

2.2 Measures

This section describes in detail the measures that we propose and evaluate. All measures yield scores between zero and one, with zero indicating a perfect match between a transcription and target phrase. For the first three measures, which are calculated programmatically, transcriptions are pre-processed to remove capitalization, punctuation, and extra spaces.

2.2.1 Lexical error rate

The traditional dictation scoring method (as described by, e.g., Buck [2001] and Irvine et al. [1974]) involves calculating the lexical error rate, which is simply the proportion of words in the target phrase that are absent in the participant's transcription. For example, for the target phrase "She wants to be a police officer," the transcription "She is a police officer" receives a score of 0.43 (3/7 of target words missing). To avoid reliance on human judgments about the source or severity of spelling errors, words must be spelled correctly to count.

2.2.2 Orthographic edit distance

The orthographic edit distance is a measure of how accurately listeners perceived the sounds of the target phrase, using letters as a proxy for sounds. Compared to the lexical error rate, it gives more credit to imperfect transcriptions containing similar sets of letters in similar orders to those of the target phrases.

We implement the orthographic edit distance between the transcribed and target phrases as the two strings' Levenshtein distance: the minimum number of single-character edits, namely, insertions, deletions, or substitutions, required to transform one into the other (Levenshtein, 1966). For instance, to transform the transcription "my fright night" into the target phrase "my Friday night" requires minimally three substitutions: replacing the last three characters of "fright." To normalize the edit distance to lie between zero and one, we divide it by the number of characters in the longer phrase, as this length represents the maximum possible distance between two items.

2.2.3 Phonological edit distance

The phonological edit distance, based on methods used to phonetically measure dialect distance (Nerbonne & Heeringa, 1997), provides a closer estimate of how well participants were able to recover the phonemes, and even the specific phonological features, of the target phrase. It is based on the same principle as the orthographic edit distance, but it uses phonemes rather than letters and captures the insight that some phonemes are more similar to each other than others. Thus, replacing a /t/ with a /d/ incurs less penalty than replacing it with /n/ because fewer features change.

To calculate the phonological edit distance, the target phrase and transcribed phrases are first converted from Latin letters to IPA characters using a word-to-phoneme dictionary, such as the CMU Pronouncing Dictionary for English (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). Words not in the dictionary, such as misspellings or uncommon names, are converted to IPA characters using a grapheme-to-phoneme engine, such as g2p_en (Park & Kim, 2018).

Once the IPA transcription of the target phrase and participant transcription are obtained, the phonological edit distance is calculated using the weighted feature edit distance of the PanPhon library (Mortensen et al., 2016), which represents every IPA segment as a vector of phonological features and weights the costs of feature edits differently depending on their class and subjective variability. To normalize the phonological edit distance to lie between zero and one, we then divide it by the weighted feature edit distance between an empty string and the longer of the two strings, as this represents the maximum possible weighted feature edit distance between them.

2.2.4 Semantic error rate

The semantic error rate gauges how well a transcription conveys the broad meaning of a target phrase. The target phrase is broken down into its key conceptual elements, defined by the phrase's open-class lemmas and personal pronouns. For example, for the target phrase "since I stopped going to the gym," the key elements are I, stop, go, and gym. We score the participant transcriptions manually by calculating the proportion of key concepts from the target phrase that are missing from the transcribed phrase, interpreting any spelling errors generously. For a noun-phrase concept to count as present, it must fill the correct thematic role in the sentence, and for a verbal concept to count, the verb's polarity (positive/negative), but not tense or aspect, has to match that of the target phrase. Thus, for the example given above, the transcription "since I'm going to the gym" receives a score of 0.25 (1/4 key concepts [stop] missing), and "since I went to Germany" scores 0.50 (2/4 key concepts [stop, gym] missing).

2.3 Methods

To evaluate the four dictation measures, we implemented them in a dictation task with reduced speech given to non-native and native listeners.

2.3.1 Participants

The participants were 116 native Dutch speakers (mean age = 21.7 years, *SD* = 2.8) with advanced L2 English proficiency and 25 native American English speakers (mean age = 24.1 years, *SD* = 2.7).

2.3.2 Materials

The dictation task comprised eight fragments of spontaneous English speech produced by a female American from Arizona in an informal dialogue. Each fragment was one or two sentences long and contained highly reduced productions. For each fragment, a critical sequence of consecutive words was selected to be the fill-in-the-blank target phrase for participants to transcribe. The target phrases and their broad phonetic transcriptions, illustrating massive reductions, are listed in Table 1.

Table 1

Dictation task target phrases

Target Phrase	Transcription of Phrase as Spoken
I didn’t really know that, but I need to take it to graduate	aɪ ɪn ɹɪli noʊ ðæ:t bət aɪ niə teɪkɪtə gɹæʤuɛt
since I stopped going to the gym	saɪ stɑp gowɪnə ʤɪm
She wants to be a police officer	ʃɑns i pəl:is ɔvəsəɹ
I was thinking of just applying to jobs in San Diego	aɪz θɪŋə ʤɪst əplaɪnə ʤɑbz ɪn sæn dieɪgoʊ
My Friday night	mʌ fɹɛ̃
she’s gonna let me know for sure today	ʃiz gənə lɛt mi noʊ fʊɹ ʃʊɹ tədeɪ
’cause that way we can be together	ksæ weɪ i kn: bi dəgɛðəɹ
I told him that I was thinking about going to	aɪ toʊld ɪm ðæt aɪz θɪŋmə goʊnə

2.3.3 Procedures

The dictation task was presented in the form of an online, self-paced Qualtrics survey with one audio fragment per page, which could be replayed as often as desired. On each page, a partial transcription of the recording was provided, and the participants’ task was to listen to the recording and to type in the missing words in the blank.

After the dictation task, all Dutch participants completed a language background questionnaire, and a subset (*n* = 45) took the LexTALE (Lemhöfer & Broersma, 2012), a measure of their English vocabulary knowledge.

2.3.4 Data pre-processing

To make the transcriptions comparable to each other and to the target phrases for automatic scoring, we processed the data so that for each contraction in the target phrases, all versions of that contraction in the transcriptions were converted to the same form (e.g., "because", "'cause", and "cuz" were all mapped onto "'cause."). As the Dutch listeners often wrote compound nouns as one word (e.g., "policeofficer" for "police officer"), we separated these forms into two words to avoid penalizing this error pattern relating to orthography rather than speech perception.

2.4 Results

The four dictation measures clearly distinguish the transcriptions of non-native and native listeners. As shown in Figure 1, the Dutch listeners performed significantly worse than the American listeners on all measures (phonological distance (*t*(415.13) = 16.58), orthographic distance (*t*(343.27) = 17.41), lexical error rate (*t*(329.99) = 16.53), and semantic error rate (*t*(297.60) = 12.73); all *p*'s < 0.001).

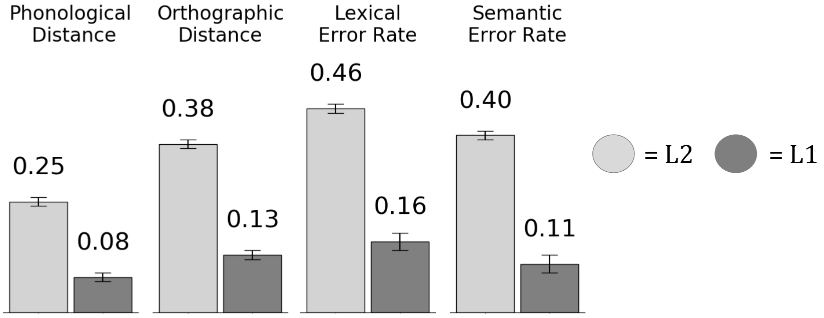

Figure 1. Mean dictation scores for the set of transcriptions made by Dutch (L2) listeners and American (L1) listeners, with bar height representing the amount of error and error bars representing the standard error of the mean

The four measures also show that participants' answers incorporate more phonological and semantic information than lexical error rate alone might suggest. Transcriptions were most different from the target phrases in lexical error rate, which was higher than orthographic distance and semantic error rate ($t(140) = 21.22$ and $t(140) = 13.25$ respectively, both p's < 0.001). Transcriptions were closest to the target phrases in phonological distance, as this score was lower than the orthographic distance, semantic error rate, and lexical error rate ($t(140) = 35.95$, $t(140) = 17.02$, and $t(140) = 34.08$, respectively, all p's < 0.001). The scores for the measures of orthographic distance and semantic error rate were equivalent ($t(140) = 1.80$, $p = 0.07$).

For each of the four measures, an overall dictation score was calculated for each participant by averaging across the eight items. Table 2 presents correlations between the Dutch listeners' four overall dictation scores and their self-rated English language proficiency in speaking, listening, reading, and writing; their average weekly hours of English listening and speaking; and their LexTALE scores.

Table 2

Correlations (Pearson's r*) between Dutch listeners' dictation scores on the four measures and language background questionnaire variables*

		PHON	ORTH	LEX	SEM
Self-Rated Proficiency	Speaking	-.19*	-.23*	-.30**	-.27*
	Listening	-.24*	-.25*	-.31**	-.32**
	Writing	-.19*	-.23*	-.27*	-.23*
	Reading	-.24*	-.29**	-.32**	-.30**
Weekly Hours	Speaking	-.03	-.09	-.07	-.11
	Listening	-.12	-.19*	-.22*	-.29**
	LexTALE	-.36*	-.35*	-.44*	-.40*

Note. * $p < 0.05$, ** $p < 0.0018$, the Bonferroni-corrected alpha

As shown in Table 3, the four measures have medium to high correlations with each other. The orthographic distance correlates highly with both the lexical error rate and phonological distance; this follows from the fact that they all depend on the specific letter sequences in the transcription for their calculation. As to be expected, the lowest correlation is between the semantic and phonological measures.

Table 3

Matrix of Pearson correlation coefficients for the transcriptions' scores on the four measures

	PHON	ORTH	LEX	SEM
PHON	1.00	.90	.77	.67
ORTH	.90	1.00	.92	.79
LEX	.77	.92	1.00	.87
SEM	.67	.79	.87	1.00

2.5 Discussion

This paper demonstrated how four measures targeting accuracy at different levels, with different degrees of granularity involved in their calculation, can be used to score dictation data, thereby increasing the amount of information that dictation tasks can yield for speech perception research.

From a practical standpoint, the easiest measures to implement are lexical error rate and orthographic edit distance as they are both calculated automatically and do not need an external data source. Phonological edit distance, while automatically calculated, requires a dictionary for converting words or graphemes to phonemes, which may be hard to find for some languages. Semantic error rate, relying on a human rater, is more time-consuming, subjective, and error-prone. It could conceivably be automated with the right language model, but the time investment may be prohibitively high except for very large data sets.

Given the four measures' high intercorrelations, using a subset of them can still be informative. For instance, the lexical and orthographic measures, both based on the degree of matching between the letter sequences in the transcription and target phrase, provide almost the same information except that the former is binary (a word matches exactly or not at all) while the latter is gradient (similarly spelled words are less penalized). Thus, unless spelling accuracy is of additional theoretical interest, the orthographic edit distance could be used by itself as it already provides a very good estimate of word recognition ability.

Combining the phonological edit distance and the semantic error rate, which themselves have a lower intercorrelation, sheds light on different aspects of performance: how accurately phonological features were recovered from the acoustic signal and how well the meaning of the utterances was comprehended. As listeners may employ different transcription strategies, prioritizing either bottom-up or top-down information, using both measures paints a more complete picture of their abilities.

Using writing as a proxy for speech perception comes with some caveats. For non-native listeners, whose sound-to-orthography mappings can differ from those of native listeners, dictation performance may be less informative about their actual sound representations. Also, since listeners tend to write real words even when they are not a perfect match for the perceived input, errors in letter sequences unrelated to the sounds actually perceived can arise. Still, the phonological distance measure allows the dictation task to evaluate phoneme perception, even for English with its notoriously irregular spelling system.

Overall, the combination of lexical, orthographic, phonological, and semantic similarity measures provides richer information than the traditional word error rate about what linguistic units listeners recover from the speech input. While we have shown how these measures can be used to analyze transcriptions of reduced speech, they are also suitable for any research on speech perception in difficult conditions, whether these involve properties of the speech itself, background noise, or listener characteristics.

Chapter 3: The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input

Abstract:
This article 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.

This chapter is based on the following:
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.

3.1 Introduction

This paper 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 doubtless shaped by experiences in conversation, as demonstrated by research on perceptual adaptation (e.g., Norris, McQueen & Cutler, 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 2017). 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, Brown et al., 1983; 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, Pickering, & Cleland, 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 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.

3.2 Ventriloquist paradigm methodology

3.2.1 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.

3.2.2 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.

3.2.3 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.

3.3 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.

3.3.1 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 followed 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.

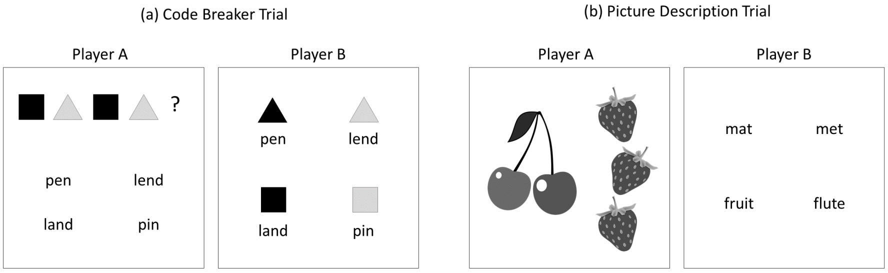

Figure 1. Sample screens for two players in one trial of the Code Breaker game (Section 3.3.1) and for one trial of the picture description game (Section 3.3.2).

3.3.2 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*").

3.3.3 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.

3.4 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 pre-recorded. 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; $\chi^2(1, N = 101) = 0.50$, $p = 0.48$.

3.5 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 3.4, 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 pre-recorded 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 ($\chi^2(1, N = 123) = 17.375, p < 0.001$); 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.

3.6 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 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 3.4, 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.

3.7 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 (Brown et al., 1983), 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 4: The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue

Abstract:

Perceptual learning of novel accents is a critical skill for second language speech perception, but little is known about the mechanisms that facilitate perceptual learning in communicative contexts. To study perceptual learning in an interactive dialogue setting while maintaining experimental control of the phonetic input, we employed an innovative experimental method incorporating pre-recorded speech into a naturalistic conversation. Using both computer-based and face-to-face dialogue settings, we investigated the effect of two types of learning mechanisms in interaction: explicit corrective feedback and implicit lexical guidance. Dutch participants played an information-gap game featuring minimal pairs with an accented English speaker whose /ɛ/ pronunciations were shifted to /ɪ/. Evidence for the vowel shift came either from corrective feedback about participants' perceptual mistakes or from onscreen lexical information that constrained their interpretation of the interlocutor's words. Corrective feedback explicitly contrasting the minimal pairs was more effective than generic feedback. Additionally, both receiving lexical guidance and exhibiting more uptake for the vowel shift improved listeners' subsequent online processing of accented words. Comparable learning effects were found in both the computer-based and face-to-face interactions, showing that our results can be generalized to a more naturalistic learning context than traditional computer-based perception training programs.

This chapter is based on the following:

Felker, E., Broersma, M. & Ernestus, M. (submitted). The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue.

4.1 Introduction

Speech perception skills are crucial for second language (L2) acquisition, not only to benefit from aural input but also to participate in conversational interaction. While listening in a native language (L1) is usually an effortless process, spoken word recognition is harder for L2 listeners due to their less accurate word segmentation, less accurate phoneme perception, and increased lexical competition arising from L1 words (Cutler, 2012). Given these speech processing difficulties inherent to L2 listening, it may be especially hard for L2 listeners to deal with the added complication of interspeaker variability, such as accent variation (Field, 2008). Unfamiliar accents can impair speech processing and comprehension both for L1 listeners (e.g., Adank et al., 2009; Clopper & Bradlow, 2008; Floccia et al., 2006; Munro, 1998) and L2 listeners (e.g., Bent & Bradlow, 2003; Escudero & Boersma, 2004; Major et al., 2005; Pinet et al., 2011). Since L2 learners often encounter multiple regional and foreign accents of their L2 in educational and professional settings abroad, the ability to adapt their perception to accommodate accent variation—what we refer to as perceptual learning—is a key communicative competency (Canagarajah, 2006; Harding 2014).

The present study aims to better understand the learning mechanisms that facilitate L2 perceptual learning of accents in dialogue. Interactive communication is considered to be an important locus of L2 acquisition in general (Ellis, 1999, 2003; Long, 1980), but perceptual learning has almost never before been researched in conversational contexts, likely due to the methodological challenge of maintaining the necessary experimental control over the phonetic input. To achieve this control to study perceptual learning in conversation, we employ an innovative paradigm in which participants interact face-to-face with a confederate whose speech is entirely pre-recorded (Felker et al., 2018). While briefly hiding her face behind a screen, the confederate uses a hidden keyboard to play pre-recorded utterances to participants' headphones. These utterances include task-relevant phrases and various flexible remarks to respond to any spontaneous questions, creating the convincing illusion of a live conversation. We also replicate the main experiment in a more traditional setup in which participants interact with what they believe to be a "smart computer player." By combining the face-to-face and computer-based settings in one study, we test whether both settings tap into the same underlying processes for L2 perceptual learning.

Within these interactive settings, we investigate two factors that may facilitate perceptual learning of the interlocutor's novel accent: (1) corrective feedback and (2) lexical guidance provided visually that constrains the interpretation of key phonemes. Perception-oriented corrective feedback and lexical guidance are two well studied mechanisms for sound learning, but neither has been studied in the context of a two-way communicative dialogue. Moreover, they are typically studied in different disciplines, with corrective feedback featuring in research on instructed L2 learning (e.g., Brown, 2016) and lexical guidance in research on L1 perception and psychophysics (e.g., Samuel & Kraljic, 2009). While L2 learning in general benefits more from explicit than implicit instruction (Norris & Ortega, 2000) and more from explicit than implicit corrective feedback (Rassaei, 2013), implicit lexical guidance may be advantageous because it retunes perception automatically (McQueen, Norris, et al., 2006) and because lexical information may have a less ambiguous interpretation than feedback in an interactive context. By studying corrective feedback and lexical guidance together, we aim to reconcile research from different fields and examine the extent to which both explicit and implicit information can contribute to L2 perceptual learning of accents.

4.1.1 Explicit perceptual learning through corrective feedback

Interactional feedback, including corrective feedback, is theorized to facilitate L2 learning because it brings learners' errors to their conscious awareness, helping them to "notice the gap" between their own productions and target forms (e.g., Schmidt, 2001). For speech perception, explicit corrective feedback in interaction would help listeners notice the discrepancy between their interpretation of the spoken input and what their interlocutor intended to communicate. Corrective feedback in the language classroom has been shown to facilitate L2 grammar, vocabulary, and pronunciation learning (e.g., see meta-analysis of Brown, 2016). However, research about corrective feedback for L2 speech perception primarily employs non-interactive settings, such as computer-based training programs using highly controlled phonetic input. These programs have been proven effective for learning L2 sounds that do not make a phonemic distinction in the L1 (e.g., Bradlow et al., 1999; Iverson et al., 2005; Wang & Munro, 2004). Furthermore, Lee and Lyster (2016b) demonstrated that the type of corrective feedback matters, using forced-choice listening tests with phonological minimal pairs. Visual corrective feedback, implemented as the word "wrong" shown onscreen, was less effective than auditory feedback, which consisted of a voice saying either “No, s/he said [X]”, “No, not [Y]”, or “No, s/he said [X], not [Y].” The most effective feedback type was the contrastive auditory feedback that combined the target and non-target forms. The authors reasoned that it was superior because it aurally reinforced the target form and increased learners’ awareness of phonetic differences by accentuating the gap between the intended forms and what they thought they heard.

To our knowledge, only one study has examined the effect of corrective feedback on L2 speech perception outside the context of computer-based training. Moving closer to a naturalistic, interactive setting, Lee and Lyster (2016a) used classroom simulations providing form-focused instruction on L2 vowel contrasts. Learners practiced their perception with pick-a-card, bingo, and fill-in-the-blank games with minimal pairs. Whenever a learner made a perceptual error, such as by selecting the wrong word in a minimal pair, the instructor repeated the learner’s wrongly chosen word verbatim with rising intonation. If the learner did not self-repair, more explicit feedback was given: “Not [Y], but [X].” Compared to a control classroom where no feedback was given, learners in the corrective-feedback classroom performed significantly better on word-identification post-tests. Taken together, Lee and Lyster’s (2016a, 2016b) studies suggest that contrastive, or more explicit, corrective feedback is more effective than generic, or implicit, feedback. It remains to be seen whether corrective feedback can also facilitate the learning of a novel accent, rather than a novel L2 phonemic contrast. Moreover, as even classroom-based corrective feedback is not always interpreted by learners the way teachers intended (Mackey et al., 2007), the interpretability of corrective feedback in a communicative dialogue merits further study. To determine whether feedback was interpreted accurately, it can be informative to examine learners’ immediate response to the feedback, or uptake (Mackey et al., 2000). The present study compares the interpretability of two types of corrective feedback in dialogue—generic and contrastive feedback—to assess which better promotes uptake for L2 perceptual learning.

4.1.2 Implicit perceptual learning through lexical guidance

Not only does interaction provide the opportunity for explicit learning through corrective feedback, but it also creates the context for implicit learning. The most-studied implicit learning mechanism for perceptual adaptation to accents is lexically guided learning (McQueen, Cutler, et al., 2006; Norris et al., 2003):

listeners use top-down lexical knowledge to constrain their interpretation of ambiguous sounds and, after exposure to those ambiguous sounds in different lexical frames, adjust their phonemic boundaries to accommodate the accent. For instance, if a speaker repeatedly pronounces /æ/ in different lexical contexts where /ɛ/ is expected (e.g., pronouncing "west" as /wæst/ instead of /wɛst/), the listener's perceptual /ɛ/ category eventually expands to allow for /æ/-like realizations. Lexical guidance provides feedback from the lexical to the prelexical level of processing (Norris et al., 2003), and the resultant perceptual learning occurs automatically as a result of exposure to ambiguous sounds in lexically-biased contexts (McQueen, Norris et al., 2006). The effects of lexical guidance on perceptual learning can be measured with a lexical decision task. For instance, Maye et al. (2008) showed that L1 English listeners who heard systematically lowered front vowels (e.g., /ɛ/ lowered to /æ/) within a short story adapted their post-test auditory lexical decision judgments in accordance with the vowel shift (e.g., becoming more likely to judge /wæb/, an accented pronunciation of "web", as being a real word). The effect of lexical guidance in perceptual adaptation is typically studied with L1 listeners and almost exclusively in non-interactive tasks due to the requirement for highly phonetically controlled stimuli (see reviews by Samuel & Kraljic, 2009; Baese-Berk, 2018). Thus, it remains to be seen how effective lexical guidance is for perceptual learning in an interactive, L2 listening context.

In recent years, lexically guided perceptual learning has also been demonstrated in L2 listening. Mitterer and McQueen (2009) showed that adding English-language subtitles to videos of heavily accented Australian or Scottish English speech improved Dutch listeners' subsequent perceptual accuracy for the dialects, supporting the theory that the lexical guidance provided by the subtitles facilitated perceptual retuning of the accented sounds. Drozdova et al. (2016) showed that Dutch listeners could adapt to an ambiguous sound between /ɹ/ and /l/ embedded into an English short story, shifting their phonemic category boundary in a different direction depending on whether they had heard the sound in /ɹ/- or /l/-biasing lexical contexts. Lexically guided learning in an L2 has been attested not only when the L2 is phonologically similar to the L1, such as with Swedish L2 listeners of German (Hanulíková & Ekström, 2017) and German L2 listeners of Dutch (Reinisch et al., 2013) and English (Schuhmann, 2014), but also when the L2 is phonologically unrelated to the L1, as with English L2 listeners of Mandarin (Cutler et al., 2018). However, crosslinguistic constraints on L2 perceptual learning have also been observed. For instance, Cooper and Bradlow (2018) showed that after exposure to accented English words presented in a lexically or semantically disambiguating context, Dutch listeners exhibited perceptual adaptation for words containing the trained accent pattern, but only for deviations involving phoneme pairs that were contrastive in both the L1 and L2.

To the best of our knowledge, lexically guided perceptual learning has never before been demonstrated in conversational interaction, where the listener also has to produce speech. Leach and Samuel (2007) found evidence that lexically guided perceptual adaptation involving newly-learned words was severely impaired when the participants had both heard and spoken the words aloud during the word training, compared to a condition in which they had only passively listened to the words during training. Baese-Berk and Samuel (2016) showed that in a feedback-based discrimination training paradigm, perceptual improvement for novel L2 sounds was disrupted when listeners had to intermittently produce speech as part of the training, possibly due to increased cognitive load. Similarly, Baese-Berk (2019) found that producing speech during training disrupted perceptual learning of novel sound categories in an implicit distributional learning paradigm, and she proposed that this may result from an overload in shared cognitive processing resources between the perception and production modalities. Overall, these studies suggest that more research is needed about the effectiveness of implicit perceptual learning mechanisms in cognitively demanding, interactive settings.

4.1.3 The present study

This study investigates the effectiveness of two types of corrective feedback and of lexical guidance on perceptual learning of a novel L2 accent in conversation. Native Dutch-speaking participants engaged in a task-based dialogue in English with an interlocutor whose accent contained an unexpected vowel shift, whereby /ɛ/ was pronounced as /ɪ/. These vowels were chosen for three reasons: (1) Dutch listeners should already perceive them as two different phonemes, given that they are also contrastive in Dutch (e.g., Booij, 1999), (2) they distinguish many English minimal pairs, facilitating the creation of experimental stimuli, and (c) this vowel shift is phonologically plausible, as short front vowel raising has been observed in various Southern Hemisphere English dialects, such as New Zealand (Kiesling, 2006; Maclagan & Hay, 2007), Australian (Cox & Palethorpe, 2008), and South African English (Bowerman, 2008). We restricted the accent manipulation to this single vowel shift and inserted it into an unfamiliar regional dialect (see Materials for details) to ensure that all participants would begin the experiment with no prior knowledge of the overall accent. To carefully control participants' phonetic exposure, all of the interlocutor's speech was pre-recorded and scripted to avoid the experimental sounds outside of critical utterances.

In each round of the interactive information-gap task, named "Code Breaker", the participant's task was to recognize a visual pattern in a sequence of shapes on their computer screen and tell their interlocutor what shape should follow to complete the sequence. The interlocutor would then tell the participant to click on one of the four words displayed on the participant's screen, as the interlocutor's screen supposedly indicated that this word was linked to the participant's shape. As the four word options always consisted of two phonological minimal pairs, the participant had to listen carefully to their interlocutor's pronunciation to choose the right word. On critical trials, the target word was spelled with "e" and contained /ɛ/ in Standard English but was pronounced with /ɪ/ instead, reflecting the experimental vowel shift.

Participants played Code Breaker in one of four conditions which differed in the mechanism available to learn the /ɛ/-/ɪ/ vowel shift. The Control condition contained no evidence for vowel shift. Whenever the interlocutor said a critical target word (e.g., "set" pronounced /sɪt/), both the "e"-spelled and "i"-spelled member of the relevant minimal pair (e.g., "set" and "sit") were among the onscreen word options. Control participants never received feedback about their selection and could thus assume their interlocutor's /ɪ/-pronounced word matched the "i"-spelled word onscreen, as it would in Standard English.

In the two corrective feedback conditions, the interlocutor responded verbally to incorrect choices. In the Generic Corrective Feedback condition, whenever the participant incorrectly selected the "i"-spelled competitor instead of the "e"-spelled target, the interlocutor simply remarked that a mistake was made (e.g., "Oh no, that's not the one!"). In the Contrastive Corrective Feedback condition, she instead used more specific phrasing that contrasted the target with the competitor (e.g., "Oh, you wanted "set" /sɪt/, not "sit" /sit/!", the /ɪ/ being pronounced /i/ to follow the vowel-raising pattern). In both conditions, we expected participants to become explicitly aware that the word they had originally understood did not match the word their partner was trying to communicate.

In the Lexical Guidance condition, evidence for the vowel shift was implicit and came exclusively from how the onscreen lexical options constrained the possible interpretation of the phonetic input. Crucially, the "e"-spelled target word was shown paired with a consonant competitor (e.g., target "set" and competitor "pet"), while the "i"-spelled option (e.g., "sit") was absent. Thus, the lexical context would imply that the /ɪ/ heard was meant to represent /ɛ/ (e.g., /sɪt/ could only match "set"), promoting lexically guided learning of the shift.

The amount of perceptual learning that occurred during the dialogue was measured in two ways. First, word identification accuracy in the critical Code Breaker trials was taken as a measure of listeners' uptake: the degree to which they correctly interpreted recent corrective feedback or lexical guidance to accommodate to the accent.[1] Second, an auditory lexical decision task following the dialogue was taken as a measure of listeners' online processing of the vowel shift. In this task, the same interlocutor produced a series of (pre-recorded) words and pseudowords, some pronounced with /ɪ/, and participants had to make speeded judgments about whether each one was a real word or not. For two critical item types, the expected response (yes/no) would differ depending on whether or not the /ɪ/ was perceived as representing /ɛ/.

A final aim of this study was to investigate perceptual learning within two different interactive settings. Accordingly, one participant group completed the entire experiment while interacting face-to-face with the experimenter, who surreptitiously played the pre-recorded utterances to participants' headphones to create the illusion of a live conversation (using the "ventriloquist paradigm"; see Chapter 3). The other participant group completed the same experiment without the interlocutor co-present; instead, they were told they were interacting with a "smart computer player." While the computer-player setting resembles traditional speech perception experiments, the face-to-face setting resembles real-life social interaction, a more typical context for L2 acquisition.

The research questions and hypotheses are as follows:

RQ1: (a) Does corrective feedback about erroneous perception of a novel accent in dialogue lead to uptake during the interaction for L2 listeners? **(b)** If so, which is more effective: generic or contrastive corrective feedback?

[1] Our definition of uptake differs slightly from what is typically used in L2 acquisition research, e.g., "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance" (Lyster & Ranta, 1997). Translating this concept to apply to how people respond to feedback about perception, rather than production, we consider making more accurate word identification responses following feedback about perception to be analogous to making verbal self-repairs following feedback about production.

H1: (a) We predict that corrective feedback about a dialogue partner's novel accent will lead to uptake and improve L2 listeners' perceptual accuracy for accented words over the course of the conversation. That is, compared to Control participants who received no evidence for the vowel shift and whose accuracy should remain close to zero, listeners in both the Generic and Contrastive Corrective Feedback conditions should show a pattern of increasing accuracy across the critical Code Breaker trials. **(b)** We further expect contrastive feedback to be more effective than generic feedback because the former is more explicit and interpretable and thereby better promotes noticing the gap between the word the listener perceived and the word the speaker intended.

RQ2: Do corrective feedback and lexical guidance about a novel accent in dialogue improve L2 listeners' subsequent online processing of the accent?

H2: We predict that corrective feedback and lexical guidance will directly improve online processing of accented speech in the auditory lexical decision task. We expect to observe the most improved processing in listeners who played Code Breaker in the Lexical Guidance condition because the lexical guidance will be automatically processed and entail less room for ambiguity in interpretation than corrective feedback during the dialogue. We also expect participants who received Contrastive Corrective Feedback to show more improved processing than those who received Generic Corrective Feedback, again because the more explicit feedback type will be more clearly interpretable. Moreover, we expect that each individual's Code Breaker accuracy itself, as a measure of their uptake for the accent, will predict their online processing even more robustly than the experimental condition in which they played the game.

In the lexical decision task, we define improved online processing as being faster and more likely to accept Critical Words with /ɪ/ pronunciations representing /ɛ/ (e.g., /bɪst/ as the pronunciation of "best"), which would sound like non-words to naïve listeners. Accepting these words would indicate that listeners have expanded their /ɛ/ category boundary to include /ɪ/-like pronunciations. We also explore whether listeners learn an even stricter rule, that /ɪ/ not only can but must represent /ɛ/, by examining whether they become more likely and faster to reject Critical Pseudowords with /ɪ/ pronunciations representing /ɛ/ (e.g., /gɪft/ as the pronunciation of "geft"). This would require overriding the real-word interpretation of these items (e.g., "gift"), reflecting an even stronger form of learning.

RQ3: Does the amount of perceptual learning of a novel L2 accent in dialogue differ between a computer-based setting and a face-to-face setting?

H3: We might expect to observe more perceptual learning in the face-to-face setting than in the computer-based setting because listeners may experience stronger social resonance with a human interlocutor. Successful perceptual learning entails listeners aligning their phonological representations to their interlocutor's, and linguistic alignment is known to be affected by social factors as well as by the perceived human or computer nature of the interlocutor (Branigan et al., 2010). On the other hand, we might find no differences in learning between the two settings, which would in any case show that results from the more traditional, computer-based setting generalize to a more naturalistic context.

4.2 Methodology

4.2.1 Participants

The participants were 108 native Dutch speakers, assigned to conditions on a rotating basis such that 27 people were tested in each of the four conditions (Control, Generic Corrective Feedback, Contrastive Corrective Feedback, and Lexical Guidance). Per condition, 15 participants were tested in the face-to-face setting and 12 in the computer-player setting; we tested more in the face-to-face setting in case we would need to exclude participants due to technical problems arising from this more complicated experimental setup (in the end, no such problems arose). Participants were aged 18 to 30 (*M* = 21.7, *SD* = 2.6) years, and 60.2% were female. All were raised monolingually and reported that English was their most proficient L2. On average, they reported speaking English 1.6 hours per week (*SD* = 2.8 hours) and listening to English for 11.9 hours per week (*SD* = 10.6 hours). On a scale ranging from 0 ("no ability") to 5 ("native-like ability"), participants' mean self-rated English proficiency was 2.9 for speaking (*SD* = 0.8), 3.3 for listening (*SD* = 0.7), 3.0 for writing (*SD* = 0.9), and 3.7 for reading (*SD* = 0.7). These measures of English usage and proficiency did not differ significantly between participants across the four conditions (all *p*'s > .05). Participants in the four conditions also reported similar levels of prior familiarity with Australian (*F*(3, 104) = 0.68, *p* = .56) and New Zealand English (*F*(3, 104) = 0.33, *p* = .80), making it unlikely that any one group would be more familiar with the experimental vowel shift. All participants gave written informed consent and received course credit or financial compensation in exchange for participating.

4.2.2 Procedures

4.2.2.1 General procedures

Participants played 84 rounds of the Code Breaker game in one of four conditions (Control, Generic Contrastive Feedback, Contrastive Corrective Feedback, or Lexical Guidance) and in one of two settings (face-to-face or computer player). Directly afterward, they completed 96 auditory lexical decision trials in the same setting. The Code Breaker game typically lasted fifteen to twenty minutes, and the lexical decision task took about six to ten minutes.

4.2.2.1.1 Face-to-face setting

The 60 participants in the face-to-face setting interacted with an interlocutor who was actually the experimenter, pretending to be just another participant, while another researcher took charge of the session. To prevent any spoken interaction between the participant and experimenter before the start of the game (which would have revealed the mismatch between the pre-recorded speech and the experimenter's own voice), the participant received instructions for Code Breaker in a separate room before the start of the main task. The researcher in charge of the session then guided the participant into the main testing room, where the experimenter was already sitting with headphones on and pretending to be busy finishing another task, staring intently at her screen and pressing buttons. The participant was quickly instructed to take a seat behind their own monitor, across the table from the interlocutor, and to put on their headphones for the remainder of the experiment. From that point on, the experimenter communicated with the participant by using a hidden keypad to play different categories of pre-recorded speech, ducking her face behind her monitor whenever "speaking" (for technical implementation details, see Chapter 3). The illusion that the pre-recorded speech was actually being spoken in real time was supported by a cover story, explained in the previous room, that both players would be speaking into microphones that transmitted their speech into each other's headphones.

4.2.2.1.1 Computer-player setting

The 48 participants in the computer-player setting received the same instructions for the Code Breaker and lexical decision tasks as the participants in the face-to-face setting except they were told that their interlocutor for both tasks was a smart computer player capable of recognizing their speech and talking back. In fact, the role of the computer player interlocutor was played by the experimenter, who listened to participants' speech from outside the testing booth via headphones. As in the face-to-face setting, she used a keyboard to control the playing of the pre-recorded utterances into participants' headphones.

4.2.2.2 Code Breaker game

Each Code Breaker trial featured a puzzle sequence and four words consisting of two minimal pairs, which appeared on the participant's and interlocutor's screens as shown in Figure 1. The participant's screen displayed the puzzle sequence above the four words, randomly positioned in four quadrants. Participants had been instructed that their partner's screen displayed four potential answer shapes, each linked to one of the four words, and that the trial's target word was linked to the correct answer shape for the puzzle.

In each trial, the participant's task was to figure out the pattern in their shape series and tell their interlocutor what shape was needed to complete the sequence (for details about the puzzles, see Section 4.2.3.1.3). The interlocutor then responded by telling the participant which word was linked to the requested shape, playing the pre-recorded target word utterance for that trial (e.g., "Okay, so you want *tab*", "That's, uh, *chase*"). Finally, the participant had to click on that word to complete their turn. No matter what shape the participant named, the interlocutor responded by playing that specific trial's target word utterance. If requested to do so, the interlocutor would repeat the word up to two times, playing additional audio tokens. Once the participant clicked on a word, it was highlighted with a gray rectangle on both players' screens, confirming the selection and ending the round. When appropriate, the interlocutor could play utterances belonging to various pre-determined categories to react to the participant's spontaneous remarks or questions; e.g., she could play affirmative responses (e.g., "Uh-huh"), negative responses (e.g., "Um, no"), statements of uncertainty (e.g., "I don't know"), reassuring remarks (e.g., "No problem!"), and backchannels to indicate listening (e.g., "Mm-hmm"). In the first few Code Breaker trials, the interlocutor would play a short affirmative utterance to indicate that she had seen the participant's choice.

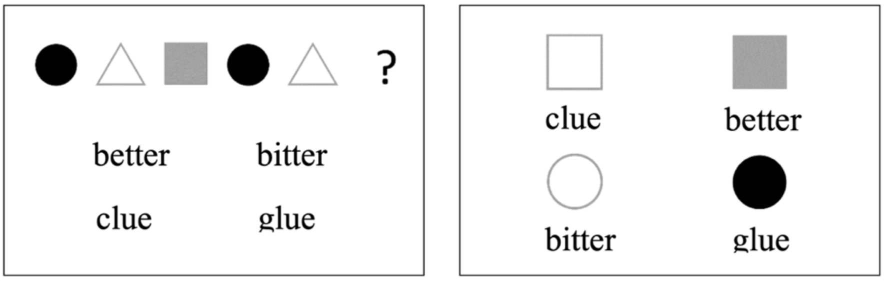

Figure 1. Example screens of a single Code Breaker trial as it appeared for the participant (left) and the interlocutor (right). Here, the puzzle's correct answer is a gray square, which corresponds to the trial's target word "better." In the Lexical Guidance condition, the phonological distractor "bitter" would be replaced with "letter" (on both screens).

4.2.2.2.1 Control condition and Lexical Guidance condition

In both the Control and Lexical Guidance conditions, participants received no feedback of any kind about whether their answers were right or wrong.

4.2.2.2.2 Generic Corrective Feedback condition

In this condition, whenever the participant clicked on a word, the word "CORRECT" in a green box or "INCORRECT" in a red box appeared in the middle of their screen as visual corrective feedback alerting them to their mistake (similar to Lee & Lyster, 2016b). If the participant had answered incorrectly, whether on a critical or filler trial, the interlocutor responded by playing a generic corrective feedback utterance (e.g., "Oh no, wrong one", "Oh no, wasn't that one"). If the participant acknowledged their error aloud before the feedback could be played, the interlocutor instead played a reassuring utterance (e.g., "That's okay!") in order to be more socially appropriate.

4.2.2.2.3 Contrastive Corrective Feedback condition

This condition worked exactly as in the Generic Corrective Feedback condition except that, on critical trials, the interlocutor reacted to errors by playing the *contrastive* corrective feedback utterance associated with that trial's target word and phonological competitor (e.g., "Oh, the answer was *set*, not *sit*", "Oh no, you wanted *better*, not *bitter*"). Generic feedback utterances would have still been played in response to errors on filler trials, but fillers (in any condition) virtually never evoked errors in practice.

4.2.2.3 Auditory lexical decision task

Instructions for the lexical decision task, presented to participants as the "Word or Not?" game, appeared onscreen right after the last Code Breaker round. The participant read that the other player was going to pronounce a series of real words and non-existing words, one at a time. Based only on the interlocutor's pronunciation, the participant had to judge whether or not each item was a real word by pressing "Y" or "N" on a button box. The audio recording of each word played automatically after a random delay of 500 to 1500 ms following trial onset, supporting the illusion that the interlocutor was reading and pronouncing the words in real time. Once the participant responded, a blank screen flashed for one second before the next trial; no feedback was provided.

4.2.3 Materials

4.2.3.1 Code Breaker game

4.2.3.1.1 Minimal word pairs

The Code Breaker game featured 16 critical minimal word pairs consisting of a target word and a phonological competitor (see Appendix A). All critical target words were spelled with "e" and contained /ɛ/ in standard English pronunciation (e.g., "set"). In the Control Condition and both Corrective Feedback conditions, each target's /ɛ/ was replaced with /ɪ/ to form the phonological competitors (e.g., target "set" paired with competitor "sit"). In the Lexical Guidance condition, the critical phonological competitors were formed by replacing one of the target word's consonants (e.g., target "set" with competitor "pet"). All critical minimal pairs consisted of mono- and disyllabic words of medium to high frequency in the SUBTLEX-UK corpus (Van Heuven et al., 2014).

The game also included 64 filler minimal word pairs, comparable to the critical pairs in length and frequency, designed to draw attention away from the critical /ɛ/-/ɪ/ contrast. To balance out the 16 critical pairs' "e"-spelled targets and "i"-spelled competitors, there were an additional 16 pairs with "i"-spelled targets (vs. competitors with any non-experimental vowel, e.g., target "bike" with competitor "bake") and 16 pairs with "e" -spelled competitors (vs. targets with any non-experimental vowel, e.g., target "tall" vs. competitor "tell"). Importantly, the "i"-spelled items in the former group were always pronounced with /aɪ/ rather than /ɪ/ to avoid providing additional information about the /ɪ/ sound, and the "e"-spelled items in the latter group, being competitors rather than targets, were never actually pronounced. Finally, 16 filler pairs had various initial-consonant contrasts (e.g., “down” vs. “town”) and 16 had final-consonant contrasts (e.g., “proof” vs. “prove”).

Each Code Breaker trial included four words: one target word and its phonological competitor plus a distractor minimal pair. To form trial lists, each critical and filler minimal pair was used once as the target pair (i.e., the pair whose target word was the right answer for that trial) and once as the distractor pair. The minimal pairs were pseudo-randomly combined into trials such that no trial combined two minimal pairs of the same contrast type. The order of the main 80 trials (16 critical + 64 filler) was pseudo-randomized such that any two trials with critical target pairs were separated by at least two trials with filler target pairs. Each trial list was then prepended with a set of four fixed word quadruplets comprising relatively easy minimal pairs as warm-up items, yielding 84 total trials.

4.2.3.1.2 Pre-recorded speech

All pre-recorded speech was scripted to avoid any instances of /ɛ/, /ɪ/, or /i/ except within the target words and contrastive corrective feedback, thereby ensuring controlled exposure to the vowel shift across conditions and preventing incidental learning of the vowel shift from the carrier phrases. The utterances were recorded at 44.1 kHz with a headset microphone in a sound-attenuating booth by a young adult female native speaker of Middlesbrough English. Her accent differed from Standard British English in several ways, e.g., /t/ was often glottalized, /ʌ/ was pronounced as /ʊ/, /eɪ/ was monophongized to /eː/, and /əʊ/ was monophongized to /ɔː/. Crucially, for the purposes of this experiment, a short front vowel shift was introduced into her accent such that she pronounced /ɛ/ as /ɪ/ and /ɪ/ as /i/. Thus, all critical /ɛ/-containing target words were pronounced with /ɪ/, and their /ɪ/-containing phonological competitors (only heard in the contrastive feedback utterances) were pronounced with /i/. This effect was achieved by replacing certain words in her script (e.g., replacing “set” with “sit” and “sit” with “seat”) and, if necessary, eliciting the desired pronunciation with pseudowords (e.g., replacing “middle” with “meedle”).

4.2.3.1.3 Puzzles

The Code Breaker game included 84 unique puzzles (see examples in Appendix A). Each puzzle was a sequence of five colored shapes followed by a question mark representing a missing sixth item, whose identity could be determined by a pattern in the preceding sequence (e.g., alternating colors or shapes). The puzzles varied in difficulty to keep the task engaging but were easily solvable within a few seconds. They were distributed randomly across trials so that the same puzzles were not always combined with the same target words (except the four puzzles fixed to the warm-up trials).

4.2.3.2 Auditory lexical decision task

The auditory lexical decision task consisted of 96 items recorded by the same speaker as in the Code Breaker game, none of which had appeared previously in the experiment (see Appendix B). There were two critical item types whose lexical status hinged on whether or not their stressed /ɪ/ vowel was interpreted as representing /ɛ/. The 12 Critical Real Words (e.g., "best") contained /ɛ/ in Standard British English but were pronounced with /ɪ/ (e.g., /bɪst/) in accordance with the vowel shift, thereby sounding like non-words (e.g., *"bist") to a naïve listener. The 12 Critical Pseudowords (e.g., *"geft") also contained /ɛ/ in Standard British English but were pronounced with /ɪ/ following the vowel shift (e.g., /gɪft/), thereby sounding like real words (e.g. "gift").

The lexical decision task included three filler item types. To draw attention away from the many /ɪ/-pronounced items, there were 36 Filler Real Words (e.g., "game") and 24 Filler Pseudowords (e.g., *"trup") that did not contain the /ɛ/, /ɪ/, or /i/ vowels. The latter were designed with the help of Keuleers and Brysbaert's (2010) software, which generated items that obeyed English phonotactic constraints and roughly matched the real words in subsyllabic structure and segment transition frequencies. In addition, there were 12 Filler /ɪ/-Pseudowords: items pronounced with /ɪ/ that would remain non-words regardless of whether or not the /ɪ/ was interpreted as /ɛ/ (e.g., /frɪp/ representing *"frep" or *"frip"). This category ensured that some of the task's /ɪ/-pronounced items had unambiguous right answers, unlike the critical items.

Overall, the lexical decision task contained 48 real words and 48 non-words. All critical and filler item groups contained a 7:5 ratio of monosyllabic to disyllabic items. The Critical and Filler Words were equivalent in their parts of speech and Zipf frequencies (Van Heuven et al., 2014). The lexical decision trial lists were ordered pseudo-randomly with two constraints: (1) at least two filler items must come between any two critical items, and (2) no streaks of five or more real words or pseudowords were allowed.

4.3 Results

For all analyses, we computed linear mixed-effects models combining data from the four conditions (Control, Generic Corrective Feedback, Contrastive Corrective Feedback, and Lexical Guidance) and both settings (face-to-face and computer player), using the *lme4* package in R (Bates et al., 2015), with *p*-values computed using Satterthwaite's degrees of freedom method of the *lmerTest* package (Kuznetsova et al., 2017).

4.3.1 Uptake during the Code Breaker game (RQ1)

First, we confirmed that corrective feedback utterances were played approximately equally often in the Generic and Contrastive Corrective Feedback conditions. The average number of critical-trial feedback utterances played per session was identical between the two conditions (*M* = 6.74 corrective feedback utterances, *SD* = 3.18 for Generic and 2.81 for Contrastive Corrective Feedback, *t*(52.22) = 0, *p* = 1), implying that any differences in uptake would likely be due to differences in the nature, rather than the quantity, of the feedback.

Table 1 presents descriptive statistics based on each participant’s overall Code Breaker accuracy, per condition, while Figure 2 displays the mean accuracy per condition over the course of the critical trials. These statistics confirm the expected near-zero accuracy in the Control Condition (which contained no evidence for the /ɛ/-to-/ɪ/ vowel shift) and the near-perfect accuracy in the Lexical Guidance condition (due to the removal of “i”-spelled competitors from the answer options).

Table 1

Overall percent accuracy on critical Code Breaker trials per participant (combining both settings, N *= 108)*

		Condition		
Statistic	Control	Generic Corrective Feedback	Contrastive Corrective Feedback	Lexical Guidance
Mean	4.4	6.0	16.2	99.1
SD	20.5	23.8	36.9	9.6
Range	0–31.3	0–56.3	0–68.8	93.8–100

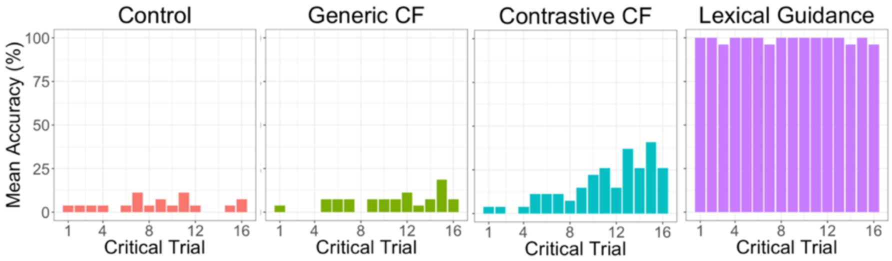

Figure 2. Mean accuracy on critical Code Breaker trials over time (combining both settings); CF = corrective feedback.

To assess whether participants adapted to their interlocutor's vowel shift during the course of the interaction—clicking on the "e"-spelled target words (e.g., "set") despite hearing /ɪ/ pronunciations (e.g., /sɪt/)—we analyzed their responses across the 16 critical trials. We computed a generalized logistic mixed-effects model with accuracy as the binary dependent variable. The fixed effects were condition (treatment coding with Control condition on the intercept), setting (treatment coding with computer player on the intercept), critical trial number (continuous variable 1–16), and all possible two- and three-way interactions among these factors. The random effects were participant and word (with random intercepts only, since random slopes prevented convergence).

The full statistical model is provided in Appendix C. The only significant simple effect in the model was an effect of condition indicating that the Lexical Guidance condition had higher overall accuracy than the Control condition (β = 11.44, *SE* = 2.52, *p* < .001, 95% CI [6.51, 16.37]). In partial support of our hypotheses, the model also contained a statistically significant interaction between condition and trial number (β = 0.34, *SE* = 0.11, *p* = .002, 95% CI [0.13, 0.55]) for the Contrastive Corrective Feedback level of condition only, indicating that these participants became more accurate on critical trials as the game went on (see the third panel of Figure 2). There were no other statistically significant interactions between trial number and either condition or setting, indicating that participants otherwise maintained a similar level of accuracy on critical trials throughout the game. Furthermore, because setting showed no significant simple or interaction effects, the face-to-face and computer-player settings appear to be equivalent.

To test the hypothesis that the two corrective feedback conditions differed from each other, we releveled the model with the Generic Corrective Feedback condition on the baseline. This releveled model revealed that the small numerical difference in accuracy between the Generic and Contrastive Corrective Feedback conditions was not significant (β = -0.63, *SE* = 1.64, *p* = .70, 95% CI [-3.84, 2.58]).

For the sake of completeness in reporting all significant effects, we also releveled the model to put Lexical Guidance on the intercept, which showed that this condition also had higher accuracy than the Generic Corrective Feedback condition (β = 12.81, *SE* = 2.57, *p* < .001, 95% CI [7.78, 17.84]) and the Contrastive Corrective Feedback condition (β = 13.44, *SE* = 2.58, *p* < .001, 95% CI [8.38, 18.51]).

In line with our predictions, Corrective Feedback participants thus showed learning over time. However, the fact that Generic Corrective Feedback participants performed no better than Control participants, showing almost no uptake for the accent, was not anticipated. Moreover, the high standard deviations and wide score ranges in both Corrective Feedback conditions (see Table 1) indicate substantial individual variability in how participants responded to the feedback.

4.3.2 Online processing in the auditory lexical decision task (RQ2)

To assess whether listeners improved their online processing of accented speech, we analyzed their responses to Critical Words and Critical Pseudowords in the auditory lexical decision task. At the outset, we removed responses with reaction times (measured from word offset) outside +/- 2 standard deviations from the mean for each item type, which amounted to 3.0% of Critical Word responses and 2.5% of Critical Pseudoword responses. In this way, we aimed to restrict the analyses to lexical decisions that were made quickly and automatically as opposed to decisions influenced by a more conscious, deliberate reasoning process.

4.3.2.1 Critical Words

Responses to Critical Words, such as “best” pronounced as /bɪst/, are summarized in Table 2 and visualized in Figure 3. Higher acceptance rates and faster reaction times to make a “yes” decision would indicate more accurate and efficient processing of the vowel shift: that listeners can (rapidly) interpret /ɪ/ as representing /ɛ/. With Control participants as the baseline, we expected to observe the most improved processing for Lexical Guidance participants, followed by Contrastive Corrective Feedback and finally Generic Corrective Feedback participants. Additionally, regardless of condition, we expected higher acceptance rates and faster "yes" reaction times for listeners who had exhibited greater uptake of the vowel shift, as measured by their Code Breaker accuracy.

Table 2

Responses to Critical Words in auditory lexical decision task

		Condition			
		Control	Generic Corrective Feedback	Contrastive Corrective Feedback	Lexical Guidance
Acceptance Rate (%)	Mean	62.3	61.3	68.2	70.3
	(SD)	(48.5)	(48.8)	(46.6)	(45.8)
Reaction Time (ms) for "Yes" Answers	Mean	675	738	664	613
	(SD)	(345)	(380)	(315)	(292)

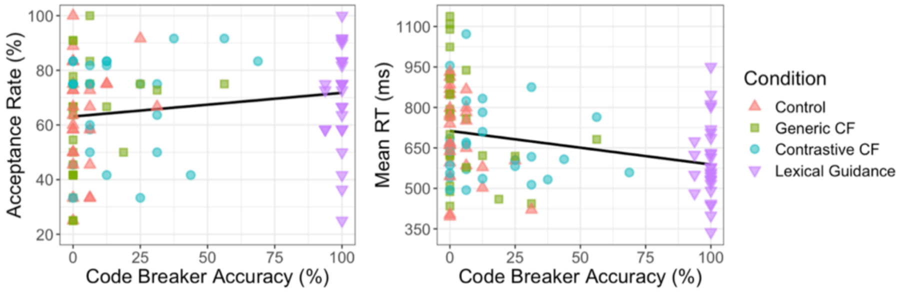

Figure 3. Critical Word acceptance rates (left) and mean reaction times for "yes" responses (right) for each participant as a function of their Code Breaker accuracy and condition, with simple regression lines; RT = reaction time, CF = corrective feedback.

4.3.2.1.1 Acceptance rates

To analyze the Critical Words' acceptance rates, we computed a generalized logistic mixed effects model with the logit link function. Response (yes/no) was the binary dependent variable. The fixed effects were condition, setting, and their interaction (all with treatment coding). The random effects were participant and item (with random intercepts only, since random slopes prevented convergence). Despite the apparent mean differences across conditions shown in Table 2, no effects proved statistically significant in the model.

We recomputed the model with condition replaced by the other predictor of interest: each participant's Code Breaker accuracy (standardized as a z-score, continuous variable). As predicted, this model showed that higher Code Breaker accuracy led to significantly more "yes" responses: β = 0.39, *SE* = 0.16, *p* = .01, 95% CI [0.08, 0.70]. There was no significant effect of setting, nor a significant interaction between setting and Code Breaker accuracy; thus, the learning effect was equivalent in the face-to-face and computer-player settings.

4.3.2.1.2 Reaction times

We restricted the reaction time analysis to trials on which participants responded "yes" to the Critical Words, computing a linear mixed effects model with log reaction time from word offset as the dependent variable and random intercepts for participant and item (no random slopes since these prevented convergence). The fixed effects included all theoretical variables of interest (condition, setting, and their interaction, with treatment coding) plus control variables known to influence lexical decision reaction times (trial number, log reaction time on previous trial, word duration, and word frequency based on the Zipf values from Van Heuven et al. [2014], all as continuous variables). Since we observed a pattern of increasing means across conditions in terms of Code Breaker accuracy (Table 1, top row) and Critical Word acceptance rates (Table 2, top row), we applied reverse Helmert coding for the condition factor, comparing each "level" of condition to the preceding levels. We applied the backward elimination procedure provided by the *lmerTest* library's step function (Kuznetsova et al., 2017) to remove insignificant predictors, resulting in a final model structure containing the significant fixed effects of condition, log reaction time on previous trial, and trial number; the final model's fit is shown in Table 3.

Table 3

Model predicting log reaction times to accept Critical Words in auditory lexical decision task

	β	*SE*	*t*-value	*p*-value	95% CI
Intercept	4.56	0.33	13.81	< 0.001*	[3.91, 5.20]
Generic CF (vs. Control)	0.08	0.06	1.31	.19	[-0.04, 0.19]
Contrastive CF (vs. Control and Generic CF)	-0.06	0.05	-1.11	.27	[-0.15, 0.04]
Lexical Guidance (vs. Control, Generic CF, and Contrastive CF)	-0.12	0.05	-2.64	.01*	[-0.22, -0.03]
Log RT of previous trial	0.28	0.05	6.06	< .001*	[0.19, 0.37]
Trial number	-0.0012	0.0005	-2.26	.02*	[-0.002, -0.0002]

Note. CF = corrective feedback, *SE* = standard error, RT = reaction time; CI = confidence interval, * = significant.

This model shows that Lexical Guidance participants were faster than those in the other three conditions to accept the Critical Words, as expected. However, no other conditions differed significantly from each other, contrary to our prediction that corrective feedback would also improve online processing. As is typically found in lexical decision experiments, reaction times were correlated with the previous trial's reaction time and became faster over time. There was no significant simple effect or interaction effect with setting after the model selection procedure, so the learning effect was comparable in the face-to-face and computer-player settings.

Next, we repeated this analysis with Code Breaker accuracy (continuous variable) replacing the condition variable. After the backward elimination procedure to remove insignificant predictors, the resulting model contained the significant fixed effects of Code Breaker accuracy, the log of the previous trial's reaction time, and trial number. As predicted, the resulting model showed that reaction times were significantly faster with increasing Code Breaker accuracy (β = -0.07, *SE* = 0.02, *p* = .002, 95% CI [-0.11, -0.03]). Furthermore, reaction times were correlated with the previous trial's reaction time (β = 0.28, *SE* = 0.05, *p* < .001, 95% CI [0.19, 0.36]) and became faster as trials went on (β = -0.0012, *SE* = 0.0005, *p* = .03, 95% CI [-0.002, -0.0002]). As the model selection procedure removed setting and its interaction with Code Breaker accuracy, it appears that reaction times in general, and perceptual learning linked to Code Breaker accuracy, were equivalent in both settings.

4.3.1.2 Critical Pseudowords

Responses to Critical Pseudowords, such as *"geft" pronounced /gɪft/, are summarized in Table 4. Recall that these pseudowords match real words in regular pronunciation (e.g., "gift"), so rejecting them requires overriding the real-word interpretation. Lower acceptance rates and faster "no" responses would indicate a very strong type of learning: that listeners (rapidly) interpret /ɪ/ as necessarily representing /ɛ/. If this type of learning were to occur, we expected to see the strongest effect (relative to Control participants) for Lexical Guidance participants, followed by Contrastive Corrective Feedback participants and finally Generic Corrective Feedback participants. Furthermore, we expected to observe a stronger learning effect in listeners who had exhibited more uptake for the accent via higher Code Breaker accuracy.

Table 4

Responses to Critical Pseudowords in auditory lexical decision task

		Condition			
		Control	Generic Corrective Feedback	Contrastive Corrective Feedback	Lexical Guidance
Acceptance Rate (%)	Mean (SD)	97.8 (14.8)	99.1 (9.7)	97.8 (14.7)	99.7 (5.6)
Reaction Time (ms) for "Yes" Answers	Mean (SD)	489 (218)	461 (196)	472 (212)	465 (220)

4.3.1.2.1 Acceptance rates

As Table 4 shows, Critical Pseudowords were almost universally accepted as real words, thereby providing no support for the strong learning hypothesis. A generalized logistic mixed effects model constructed the same way as for Critical

Words confirmed that no effects of condition or setting were statistically significant, nor were any effects significant when recomputing the model to replace the condition variable with Code Breaker accuracy.

4.3.1.2.2 Reaction times

Given the extremely high acceptance rates, there was insufficient data to analyze reaction times to "no" responses as planned. Therefore, we analyzed reaction times to "yes" responses instead, using the same model structure and selection procedure as with Critical Words. For word frequency, we used frequencies of the real words the pseudowords sounded like (e.g., the frequency of "gift" for *"geft" pronounced /gɪft/). Neither condition, setting, nor their interaction were significant predictors, leaving only control predictors in the model as shown in Table 5. When we repeated the modeling procedure with Code Breaker accuracy replacing condition, Code Breaker accuracy was also not significant, yielding an identical final model. In short, the time it took participants to accept the Critical Pseudowords was the same regardless of condition, Code Breaker accuracy, and setting.

Table 5

Model predicting log reaction times to accept Critical Pseudowords in auditory lexical decision task

	β	*SE*	*t*-value	*p*-value	95% CI
Intercept	3.94	0.46	8.64	< 0.001*	[3.05, 4.84]
Word frequency	0.21	0.08	2.73	.02*	[0.06, 0.35]
Word duration	-0.001	-0.0004	-2.48	.04*	[-0.002, 0.0002]
Log RT of previous trial	0.23	0.03	7.03	<.001*	[0.16, 0.29]

Note. *SE* = standard error, RT = reaction time; * = significant.

4.3.2.3 Filler items

Overall, the responses were as expected, with a majority of “yes” responses to Filler Words (mean = 94.4%, *SD* = 23.0%) and “no” responses to Filler Pseudowords and Filler /ɪ/-Pseudowords (mean = 78.7%, *SD* = 40.9% and mean = 86.9%, *SD* = 33.7%, respectively). See Appendix D for descriptive statistics and supplementary analyses.

4.3.3 Differences between computer-based and face-to-face settings (RQ3)

As described in the preceding sections, the setting (computer player vs. face-to-face) did not interact significantly with condition for predicting uptake in the Code Breaker game, nor did it interact significantly with either condition or Code Breaker accuracy for predicting online processing in the lexical decision task. Therefore, these results do not provide any evidence that the perceptual learning under study differs between the computer-based and face-to-face settings.

4.4 Discussion

The purpose of the present research was to investigate the effectiveness of two types of corrective feedback and of lexical guidance at improving the perceptual processing of an unfamiliar accent in an interactive, L2 listening context. To assess whether generic or contrastive corrective feedback would better promote uptake for the accent, we analyzed listeners’ word identification accuracy over the course of the interactive Code Breaker game. Furthermore, using an auditory lexical decision task, we examined whether listeners’ online processing of the accent improved, either as a direct result of receiving corrective feedback or lexical guidance or as a result of how much uptake they had exhibited during the game. Finally, we examined whether perceptual learning differed between a computer-based and a face-to-face interactive setting.

4.4.1 Comparing uptake from generic and contrastive corrective feedback

The first research question was whether corrective feedback would promote uptake by increasing word identification accuracy over the course of the interaction and, if so, whether generic or contrastive feedback would be more effective. Results showed that listeners in the Contrastive Corrective Feedback condition, but not listeners in the Generic Corrective Feedback condition, were more accurate than those in the Control condition for later-occurring critical trials in the Code Breaker game. That is, listeners receiving contrastive corrective feedback began to accommodate their interlocutor's accent over time, allowing her /ɪ/ pronunciation to represent /ɛ/ and choosing "e" -spelled words (e.g., "set") despite hearing /ɪ/-containing pronunciations (e.g., /sɪt/). The superiority of contrastive corrective feedback matches our hypothesis and mirrors the findings of Lee and Lyster (2016b), who found that the most effective feedback type for learning a non-native L2 sound contrast was one that auditorily contrasted two members of a minimal pair. One reason for the effectiveness of this feedback type could be that it drew listeners' attention to the relevant phonological contrast, simultaneously providing positive evidence for the correct interpretation and negative evidence against the wrong interpretation. Another explanation is that the Contrastive Feedback condition is the only one that provided additional exposure to the target form with each instance of feedback. However, it seems unlikely that mere exposure to a cross-category vowel shift would induce learning by itself in the absence of some sort of disambiguating information, as the lack of learning in the Control condition attests.

Although the Contrastive Corrective Feedback did improve word identification during the interaction, some listeners never accommodated the vowel shift despite the feedback. Also, contrary to our expectation, almost no listeners in the Generic Corrective Feedback condition demonstrated any uptake. This calls to mind Mackey et al.'s (2000) point that the way L2 learners perceive interactional feedback is not always in line with what their dialogue partner intended to communicate. In our case, listeners may not have interpreted the generic corrective feedback as reflecting their mistaken perception. Rather, they might have assumed that they heard right but that their partner had misspoken, or that they were told the wrong word because they gave the wrong answer to the puzzle. Moreover, repeated provision of generic feedback was arguably unnatural from a pragmatic standpoint because a cooperative interlocutor would make their remarks more specific over time or perhaps even adapt their own pronunciation in order to avoid repeated misunderstandings. While such ambiguities about feedback would not arise in a form-focused perception training program, they may occur often in interactive communication. Interestingly, Lee and Lyster's classroom-based study (2016a) implemented a two-step feedback protocol, first providing implicit feedback (repeating the wrong word with question intonation) and following it up when necessary (if learners did not make self-repairs) with explicit feedback similar to our study's contrastive corrective feedback. The fact that their learners did not always seem to understand the initial implicit feedback aligns with our finding that generic feedback about speech perception may, in some contexts, be too ambiguous to learn from.

4.4.2 Effect of feedback and lexical guidance on lexical processing

The second research question was whether corrective feedback and lexical guidance, or uptake resulting from these factors, would contribute to faster and more accurate online processing of the accented speech, as measured by a lexical decision task. Results showed that for critical accented words (e.g., "best" pronounced as /bɪst/), online processing was faster for participants receiving lexical guidance than for participants in the other three conditions, as evidenced by faster reaction times to accept these items as real words. Additionally, lexical processing was both faster and more accurate (in terms of acceptance rates) for listeners who had exhibited greater uptake, as operationalized by their word identification accuracy during critical Code Breaker trials. These results align with those of Maye et al. (2008), who found that accented words that originally sounded like non-words came to be more often interpreted as real words after exposure to a vowel shift in a story context. In our study, the fact that the online processing of accented words was more robustly affected by prior uptake, rather than being directly affected by condition, suggests that listeners' conscious word recognition during Code Breaker played a crucial role in automatizing their knowledge of the vowel shift. In other words, online lexical processing changed only to the extent that listeners had interpreted the accented words correctly during the previous communicative task.

Interestingly, the lexical decision task showed no significant effects of condition or uptake for Critical Pseudowords (e.g., *"geft" pronounced as /gɪft/), which were in fact nearly universally accepted as words by all participants (e.g., /gɪft/ was treated as "gift"). Thus, even if listeners had learned that /ɪ/ could represent /ɛ/, they did not learn that it must represent /ɛ/. This lack of learning effect for items that sounded like real words is also consistent with the results of Maye et al. (2008). They found that items that were perceived as real words before exposure to a novel accent were still judged as real words after exposure, even when the exposure had contained evidence that the vowel was involved in a chain shift (e.g., /wɪtʃ/ or "witch" was perceived as a real word both before and after accent exposure, even though the /i/-to-/ɪ/ shift in the exposure implied that /wɪtʃ/ should correspond to the non-word "weech"). Overall, our results indicate that even if listeners did adapt to the vowel shift, they did not completely remap their vowel space but simply increased their tolerance for non-standard pronunciations (i.e., allowing /ɪ/-like pronunciations of /ɛ/).

4.4.3 The interactive context for L2 sound learning

The third research question was whether perceptual learning would differ between the two communicative settings: interacting with a computer player (resembling traditional lab-based phonetic training studies) and interacting with a face-to-face interlocutor (resembling naturalistic interaction). Across all results, no significant differences in perceptual learning between the settings were observed. While we cannot draw strong conclusions from the lack of a difference, especially given the modest effect sizes, this does suggest that the perceptual learning mechanisms under study can be generalized to a more natural communicative context than what is traditionally studied in the field of L2 perceptual learning.

The fact that lexical guidance in interactive conversation improved online perceptual processing shows that this type of implicit perceptual learning can occur even when cognitive processing demands are relatively high. Not only did participants have to solve puzzles on every turn, they also engaged their L2 speech production system repeatedly to communicate the answers. While previous research found that alternating speaking and listening could interfere with perceptual learning (Baese-Berk 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007), the present findings show that significant learning can still take place in such interactive conditions. Our study was not specifically designed to test the effect of cognitive load on perceptual learning. However, other researchers have found that speaking with a physically co-present interlocutor involves a higher cognitive processing load than speaking in response to pre-recorded utterances (Sjerps et al., 2020). Thus, our experiment's face-to-face setting might well have induced a higher processing load than the computer-player setting, yet still it led to equivalent perceptual learning. Finding comparable learning effects in two different interactive settings, even in a task with relatively high processing demands, supports the viewpoint that conversational interaction is a beneficial context for L2 learning (Ellis 1999, 2003; Long 1980, 1996); moreover, we have now extended interactionist research to the area of L2 speech perception.

One important issue raised by our study is the role of explicit and implicit learning in perceptual adaptation to a novel L2 accent, as the contribution of these two types of learning to L2 acquisition is a question of interest to the field (Hulstijn, 2005). The present findings suggest that what matters for perceptual learning, whether it occurs explicitly or implicitly, is the extent to which the listener reaches the right interpretation of the spoken words during the learning phase. In our Lexical Guidance condition, the onscreen text was a reliable cue to the proper interpretation of the interlocutor's word. Before the interlocutor spoke (e.g., saying /lɪft/), participants could already rule out the incorrect default interpretation (e.g. "lift") because it was absent from their onscreen answer options; this made it easy to choose the right word (e.g. "left") despite the accented vowel. This high word identification accuracy during the interaction, confirming that participants mapped the ambiguous pronunciations to the correct lexical items, was linked to improved online processing in the subsequent lexical decision task. As lexically guided perceptual learning is an automatic process (McQueen, Norris, et al., 2006), it appears that mere exposure to vowel shift in the context of the disambiguating lexical information was enough to trigger perceptual adjustments, without interpretational difficulties playing a role. In the Corrective Feedback conditions, however, listeners were much less likely to interpret their partner's words correctly during the interaction, even after receiving repeated negative feedback in response to their perceptual errors. Thus, the corrective feedback, especially the generic feedback, was apparently not a reliable cue to interpreting the interlocutor's accent. This finding mirrors classroom-based studies (e.g., Mackey et al., 2000, 2007) showing that the linguistic target of corrective feedback is not always perceived by learners. Overall, our study suggests that the cues that drive implicit perceptual learning, like lexical constraints, may sometimes be more effective than the cues used in explicit perceptual learning, like corrective feedback, if the implicit cues yield a more reliable interpretation. Having to consciously process interactional feedback creates more room for ambiguity in interpretation due to any number of social and pragmatic factors, especially when the feedback is relatively generic in form.

Furthermore, our results suggest that conscious awareness is beneficial for L2 sound learning, supporting a weak version of the noticing hypothesis (Schmidt, 2001). Although listeners in the Lexical Guidance condition only had implicit cues to learn from, they very likely noticed that their interlocutor had a non-standard accent that they needed to adapt to, as the Code Breaker game encouraged them to choose words that mismatched their default interpretation (e.g., choosing "left" when hearing /lɪft/). Moreover, for corrective feedback to be effective, it was crucial that it explicitly highlighted the gap between the speaker's intended word and what the participant had mistakenly perceived. The positive relationship between listeners' uptake and their subsequent online processing implies that listeners who noticed these mismatches, interpreted them as reflecting their own mistaken perception, and adjusted their responses accordingly were the ones who subsequently became faster and more accurate at processing accented words.

4.4.4 Future directions and conclusions

The limitations of the present study suggest several interesting avenues for future research. First, this study only examined short-term perceptual learning based on a relatively brief dialogue with a single accented speaker. A single interactive session may not have been sufficient to produce robust learning, particularly since the amount of feedback that could be given was limited by pragmatic and methodological considerations. Thus, the potential for more robust learning effects from repeated or prolonged interaction merits further study. Moreover, while we limited the pre-recorded speech to that of a single speaker to maximize phonetic control, it would be useful to test how well the present results generalize to other voices and accents. This study focused on learning a novel L2 accent involving familiar vowels, but future research should also examine how interaction facilitates perceptual learning of L2 phonetic contrasts not present in the L1 phonemic repertoire. Additionally, given the substantial individual variability we observed in listeners' receptivity to corrective feedback, it would be interesting for future research to investigate whether factors such as proficiency can explain L2 learners' variable success in perceptual learning from interaction.

In conclusion, the main finding of this study is that L2 listeners can use corrective feedback and lexical guidance in conversation to perceptually adapt to a vowel shift in an unfamiliar accent, improving both their word identification accuracy and their online processing of accented words. Specifically, L2 listeners' word identification accuracy was shown to improve over the course of the interaction when their dialogue partner responded to perceptual errors with corrective feedback that explicitly contrasted the perceived word and the intended word. Their accuracy did not improve if they only received generic feedback, highlighting the importance of clear interpretability for interactional feedback to effectively promote uptake. The study also demonstrated that after the dialogue, L2 listeners' online processing of accented words was faster if, during the dialogue, onscreen lexical information had implicitly constrained their interpretation of the interlocutor's words. Moreover, individual differences in the amount of uptake for the accent during the dialogue significantly predicted both the speed and accuracy of post-dialogue lexical processing. Finally, as our phonetically controlled experimental paradigm yielded comparable learning effects in both computer-based and face-to-face interactive settings, these results can likely be generalized to a more naturalistic L2 acquisition context.

Chapter 5: Lexically guided perceptual learning of a vowel shift in an interactive L2 listening context

Abstract:
Lexically guided perceptual learning has traditionally been studied with ambiguous consonant sounds to which native listeners are exposed in a purely receptive listening context. To extend previous research, we investigate whether lexically guided learning applies to a vowel shift encountered by non-native listeners in an interactive dialogue. Dutch participants played a two-player game in English in either a control condition, which contained no evidence for a vowel shift, or a lexically constraining condition, in which onscreen lexical information required them to re-interpret their interlocutor's /ɪ/ pronunciations as representing /ɛ/. A phonetic categorization pre-test and post-test were used to assess whether the game shifted listeners' phonemic boundaries such that more of the /ɛ/-/ɪ/ continuum came to be perceived as /ɛ/. Both listener groups showed an overall post-test shift toward /ɪ/, suggesting that vowel perception may be sensitive to directional biases related to properties of the speaker's vowel space. Importantly, listeners in the lexically constraining condition made relatively more post-test /ɛ/ responses than the control group, thereby exhibiting an effect of lexically guided adaptation. The results thus demonstrate that non-native listeners can adjust their phonemic boundaries on the basis of lexical information to accommodate a vowel shift learned in interactive conversation.

This chapter is based on the following:
Felker, E., Ernestus, M. & Broersma, M. (2019). Lexically guided perceptual learning of a vowel shift in an interactive L2 listening context. In *Proceedings of the 20th Interspeech* (pp. 3123–3127).

5.1 Introduction

Given the inherent variability in the acoustic realization of speech sounds, both within and across speakers and dialects, the speech perception system needs to be able to adjust phonemic boundaries dynamically in order to make speech input interpretable. It has been shown that listeners can make use of high-level information from the lexicon to modify their phonemic boundaries in a process known as lexically guided perceptual learning (Norris et al., 2003). For example, when an ambiguous fricative /?/ in the spectrum between /f/ and /s/ is repeatedly substituted for word-final /f/ sounds, e.g., replacing /bəˈlif/ (from *belief*) with /bəˈli?/, listeners are more likely to label the ambiguous sound as /f/ in a subsequent phonetic categorization task, but if the same ambiguous fricative replaces word-final /s/ sounds, e.g., replacing /ˈnotəs/ (from *notice*) with /ˈnotə?/, listeners tend to subsequently categorize the sound as /s/ (Reinisch & Holt, 2014).

To date, most studies on lexically guided perceptual learning focused on ambiguous consonant sounds presented to native listeners in receptive tasks such as passive listening or lexical decision paradigms (see reviews of Samuel & Kraljic (2009) and Baese-Berk (2018)). However, the extent to which this type of learning applies more generally to other classes of speech sounds and in more cognitively demanding listening conditions remains an open question. We present an experiment that investigates whether lexical information can also retune phonemic boundaries for vowel perception in non-native (L2) listening during a task-based dialogue, thereby extending previous research to a different class of speech sounds, a lower-proficiency listener group, and a more naturalistic communicative setting.

Vowels are an interesting test case for lexically guided learning because differences in vowel sounds distinguish many dialects (e.g., Thomas, 2001), making adaptation to vowel variation crucial for communication. Despite this, few studies have specifically tested lexically driven adaptation to vowels. It has been shown that Dutch listeners can use lexical information to retune their perception of an ambiguous Dutch vowel (McQueen & Mitterer, 2005), though the learning effects in a phonetic categorization task were highly sensitive to the presentation order of various testing blocks. As for a full vowel shift, one study showed that English listeners exposed to 20 minutes of synthesized speech with systematic front vowel lowering adapted their lexical decision judgments in accordance with the vowel change (Maye et al., 2008). Another study showed that both Dutch and native English listeners adapted to a series of lowered front vowel shifts heard in 72 training items in a manipulated English accent (Cooper & Bradlow, 2018). Whether adaptation to a vowel shift can also occur with more limited exposure than in Maye et al.'s (2008) and Cooper and Bradlow's (2018) studies, given the relative instability of the perceptual adaptation in McQueen and Mitterer's (2005) study, remains to be seen. In theory, we expect a vowel shift to be learnable on the basis of lexical information, but the fact that vowel perception is less categorical than consonant perception may make the observable adaptation effect more subtle.

Relatively little research has studied lexically driven perceptual adaptation in L2 listeners, though it has been demonstrated for Dutch listeners in English with an ambiguous sound between English /l/ and /ɹ/ (Drozdova et al., 2016). Lexically driven perceptual adaptation may be generally more difficult for L2 listeners for numerous reasons: not only because incomplete L2 vocabulary knowledge may lead to differently balanced patterns of lexical activation than for native listeners, but also because increased lexical competition arises from words in their native language during word recognition (Lecumberri et al., 2010). In such circumstances, relying on lexical information to disambiguate competing interpretations of a speech sound may be relatively ineffectual. Therefore, while we predict that lexically guided learning will be possible for L2 listeners, the effect may be smaller than what has typically been shown for native listeners.

To our knowledge, the phenomenon of lexically guided perceptual learning has never before been studied in a conversational context. On the one hand, we might expect any kind of perceptual adaptation, including adaptation driven by lexical information, to occur in task-based interaction even more readily than in passive listening since listeners engaged in dialogue may be more motivated to understand their interlocutor and may therefore expend more conscious effort to comprehend an unfamiliar accent. On the other hand, conversational interaction may be more cognitively demanding as it engages the speech production system; moreover, recent evidence suggests that producing speech during perception training can interfere with learning new sound representations (Baese-Berk & Samuel, 2016). Therefore, to show that lexically guided perceptual learning still occurs in an interactive task-based dialogue, when exposure to accented speech input is repeatedly interrupted by listeners' own speech production processes, would further attest to the robustness of this type of learning.

In the present experiment, native Dutch-speaking participants played an interactive puzzle-solving computer game called "Code Breaker" with another player whose speech was pre-recorded in order to control phonetic exposure (see Chapter 3). The pre-recorded speech, belonging to a native English speaker, exhibited an unexpected vowel shift, such that /ɛ/ was pronounced as /ɪ/. As both sounds are part of the Dutch listeners' native phonetic inventory, we expect the vowel shift to be salient. The interactive game was an information gap task that required participants to alternate between solving pattern recognition puzzles aloud and then following their partner's oral instructions to click on certain members of phonological minimal pairs displayed onscreen, the target words being linked to the puzzle shapes. As previous research has shown that as few as ten training items can trigger perceptual learning (e.g., Poellmann et al., 2011), in this experiment we limited the evidence for the vowel shift to a small number of critical target words containing the vowel shift (e.g., "lesson" pronounced as /lɪsən/).

In the lexically constraining condition, the relevant minimal pair contained the target word and a competitor that differed only in a consonant sound (e.g., "lesson" and "lemon"), such that the target word remained the most plausible match to the phonetic input despite the vowel mismatch, thereby promoting perceptual learning. In the control condition, the very same target words were instead displayed alongside an /ɪ/-containing member of the minimal pair (e.g., "listen"), such that the competitor was a perfect match for the phonetic input. In the absence of any corrective feedback about their responses, participants could simply choose the /ɪ/-containing competitor, and no perceptual adaptation to their partner was necessary. A phonetic categorization pre-test and post-test, featuring vowels along a 12-step continuum between the same speaker's /ɛ/ and /ɪ/, was used to evaluate how participants' phonemic boundaries shifted as a result of their experience in the interaction.

5.2 Method

5.2.1 Participants

Thirty native Dutch speakers (8 male) aged 18 to 28 years (mean = 20.7, *SD* = 2.8) participated in exchange for course credit or financial compensation. They were raised monolingually and had intermediate to advanced L2 English proficiency.

5.2.2 Materials

5.2.2.1 Phonetic categorization in pre-test and post-test

A female native speaker of Middlesbrough English recorded the pseudowords /fɛf/ and /fɪf/, from which the /ɛ/ and /ɪ/ vowels were extracted (/ɛ/: F1 = 734 Hz, F2 = 2036 Hz; /ɪ/: F1 = 557 Hz, F2 = 2186 Hz). In Praat (Boersma, 2001), a 12-step continuum (v1 … v12) was created between the two endpoints with source-filter vowel resynthesis in which F1, F2, and F3 varied in evenly spaced steps along the continuum (F3 was allowed to vary as it improved the sound quality). The duration of all resynthesized vowels was set to 164 ms, as in the speaker's original /ɛ/.

The speaker also recorded 9 consonant-group carrier frames (/f_pt/, /f_sk/, /p_f/, /p_ft/, /sp_f/, /sp_p/, /sp_ʃ/, /θ_ ʃ/, and /t_ ʃ/), each of which was pronounced in 3 versions: surrounding the /ɛ/ vowel (e.g., /fɛpt/), surrounding the /ɪ/ vowel (e.g., /fɪpt/), and preceded by the stressed syllable /pɒp/ and surrounding a schwa (e.g., /'pɒp.fəpt/). All frames were phonotactically legal pseudowords, whether surrounding /ɛ/, /ɪ/, or /ə/. The vowels were then removed from the frames, and the 12 resynthesized vowels of the continuum were spliced into the 9 frames such that for each combination of vowel step and consonant group, 2 of the 3 frame versions were used as carriers, resulting in 216 (12 x 9 x 2) total items. Which 2 of the 3 frame versions were used was systematically shifted throughout the stimuli in a counterbalanced manner (e.g., v1 was spliced into the /f_pt/ frames made from /fɛpt/ and /'pɒp.fəpt/, v2 was spliced into the /f_pt/ frames made from /fɪpt/ and /fɛpt/, and v3 was sliced into the /f_pt/ frames made from /'pɒp.fəpt/ and /fɪpt/). As a result, each of the 9 consonant groups and 3 frame versions occurred the same number of times for each step on the continuum.

5.2.2.2 Code Breaker interactive game

The Code Breaker game featured 16 critical target words, each spelled with "e" and featuring /ɛ/ in standard English pronunciation. In the lexically constraining condition, the phonological competitor for each target word formed a minimal pair with the target by differing in one consonant sound (e.g., target "set" with competitor "pet"). In the control condition, the competitor differed from target in one vowel by replacing the /ɛ/ with an /ɪ/ (e.g., target "set" with competitor "sit"). In addition to the critical minimal pairs, both conditions contained 64 filler minimal pairs comparable to the critical pairs in word length and frequency and exhibiting various other contrasts: 16 "i"-spelled targets (pronounced with /aɪ/)

vs. non-“e”-spelled competitors, 16 non-“i”-spelled targets vs. “e”-spelled competitors, and 32 consonant minimal pairs (16 word-initial and 16 word-final differences) with non-experimental vowels.

Each Code Breaker trial comprised four words: one target word and its phonological competitor (the “foreground” minimal pair) and two unrelated competitors (the “background” minimal pair). Throughout the game, each minimal pair appeared once as the foreground pair and once as the background pair. Fifteen pseudo-randomized stimuli lists were generated, each with different combinations of foreground and background pairs and with trial orders pseudo-randomized such that any trials with critical foreground pairs were spaced at least two trials apart. Four additional fixed word quadruplets were appended to the start of every stimuli list as a practice block such that each list contained 84 trials in total (16 critical trials + 64 filler trials + 4 practice trials). Each list was used once in the lexically constraining condition and once in the control condition, the only difference between the two list versions being the critical minimal pairs.

In addition to the word quadruplets, the Code Breaker game included 84 unique puzzles, one for each trial. Each puzzle was a sequence of five shapes followed by a question mark in place of a missing sixth shape, whose identity could be determined by a pattern in the preceding sequence (e.g., alternating colors). Four puzzles were used for the practice block. The other 80 puzzles were distributed randomly across the trials in each list.

All scripted, pre-recorded speech for the Code Breaker game was recorded by the same speaker as in the phonetic categorization task. Crucially, a short front vowel shift was introduced in her accent such that the /ɛ/ vowel was pronounced /ɪ/, entailing that all critical target words were pronounced with /ɪ/. This effect was achieved by replacing the target /ɛ/-words in the speaker’s script with their phonological /ɪ/-competitors.

The pre-recorded utterances included, for each target word, a multi-word instruction telling the participant which word to click on, sometimes with a disfluency to make the speech sound more natural (e.g., “That’s, uh, *chase*”, “Okay, so you want *tab*”). There were also several categories of utterances that could be played at any time to react to participants’ questions, including affirmative and negative responses, non-lexical backchanneling to indicate listening, reassuring remarks, and task-related phrases. No scripted phrases contained any words with the /ɛ/ or /ɪ/ vowels, whether in their standard or accented pronunciation, to ensure that the only evidence for the vowel shift was the word options displayed onscreen in the lexically constraining condition.

Words with /i/ were also excluded from the speech stimuli in order to leave open the possible interpretation from the listener's perspective that the /ɛ/-to-/ɪ/ shift was part of a chain shift rather than a vowel merger.

5.2.3 Procedures

At the start of the experimental session, participants were told they would be playing two games with a smart computer player that could verbally interact with them. Participants sat in a separate testing booth so they would not notice that the experimenter was controlling the computer player's speech. The session began with the Code Breaker practice block, followed by the phonetic categorization pre-test, the main Code Breaker game, and the phonetic categorization post-test.

5.2.3.1 Phonetic categorization in pre-test and post-test

In this task, participants had to decide which of two pseudowords the computer player was pronouncing in each trial. At the start of each trial, one audio stimulus was played through a set of headphones. The item's ɛ-representation (e.g., "poptesh") was spelled out on the left side of the screen and its ɪ-representation (e.g., "poptish") on the ride side. Once the participant made a choice by pressing either the left or right button of a button box, the next trial began after a randomly determined interval of 450 to 650 ms. The same 216 stimuli were played in a different randomized order for the pre-test and post-test according to 15 trial lists, each of which was used for two participants: one in the control condition and one in the lexically constraining condition.

5.2.3.2 Code Breaker interactive game

In each Code Breaker trial, the participant's screen displayed the puzzle sequence above the set of four words. The correct answer to the puzzle and three distractor shapes appeared on the experimenter's screen, with each shape displayed above one of the same four words. The correct-answer shape always appeared together with the target word for that trial.

In each trial, participants' first task was to figure out the pattern in their sequence of shapes and to state what sixth shape would be needed to complete the sequence. In response, the experimenter played a pre-recorded utterance telling the participant which word to click on; this was always the trial's target word, regardless of what shape the participant asked for. Using different numeric keys linked to different categories of utterance types, the experimenter could play pre-recorded phrases as needed to respond to requests for help, repetition, or clarification, thereby making the game more interactive.

5.3 Results

5.3.1 Code Breaker interactive game

As expected, participants almost always clicked on the critical target word in the lexically constraining condition, despite the vowel mismatch (mean target word responses = 98.8%, *SD* = 11.1%), while they almost never chose the critical target word in the control condition (mean target word responses = 4.6%, *SD* = 21.0%). This difference between conditions was significant: $t(363.96) = 61.48$, $p < 0.001$. Thus, the game effectively caused listeners in the lexically constraining condition to actively choose /ɛ/-words (e.g., “lesson”) when hearing /ɪ/ pronunciations (e.g., /lɪsən/) from their partner.

5.3.2 Phonetic categorization in pre-test and post-test

To assess whether the lexically constraining condition led to a shift in phonemic boundaries between /ɛ/ and /ɪ/, we analyzed participants’ responses along the 12-step vowel continuum in the phonetic categorization pre-test and post-test using mixed-effects logistic regression models with the binomial link function in the lme4 package in R (Bates et al., 2015). Response was the binary dependent variable, participant and consonant frame were random effects, and test time (pre vs. post), condition (control vs. lexically constraining), and vowel step (continuous 1–12) were fixed effects; no random slopes were included in the final model due to lack of convergence.

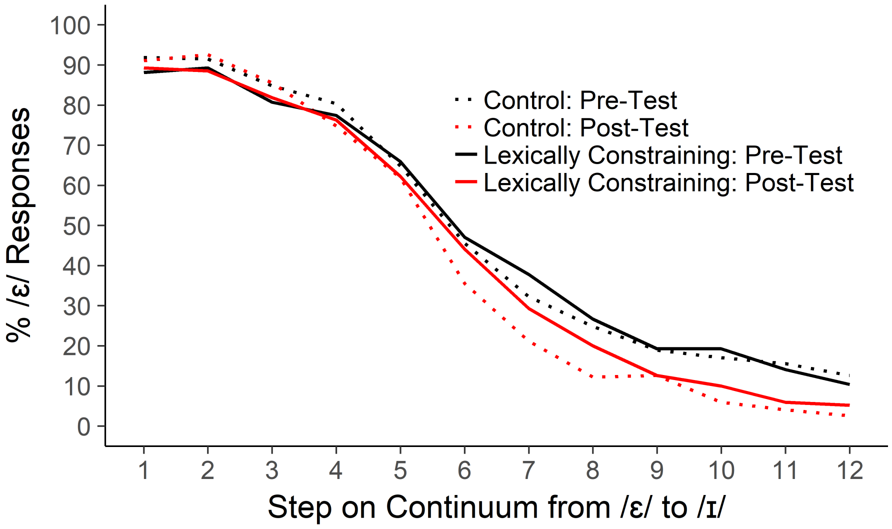

Figure 1. Phonetic categorization responses.

The mean responses are shown in Figure 1. The proportion of /ɛ/ responses was significantly higher toward the /ɛ/ end of the continuum (β = -0.74, *SE* = 0.02, *p* < 0.001), but contrary to our expectation, fewer /ɛ/ responses were made in the post-test than the pre-test for both groups (β = -0.68, *SE* = 0.18, *p* <0.001). There were significant interactions between vowel step and test time (β = 0.20, *SE* = 0.03, *p* < 0.001) and between vowel step and condition (β = 0.08, *SE* = 0.03, *p* < 0.05); these indicate that the shift toward /ɪ/ responses in the post-test, as well as the difference between conditions, was greater for vowels closer to the /ɪ/ end of the continuum. The three-way interaction between vowel step, test time, and condition was significant (β = -0.08, *SE* = 0.04, *p* < 0.05). Thus, the tendency to shift toward /ɪ/ responses in the post-test for vowels on the /ɪ/ end of the continuum was reduced in the lexically constraining condition relative to the control condition; in other words, listeners in the experimental group were more likely than control-group listeners to label the /ɪ/-like items as /ɛ/ items in the post-test. To clarify the overall effect, Figure 2 illustrates the mean response data collapsed across all steps of the vowel continuum.

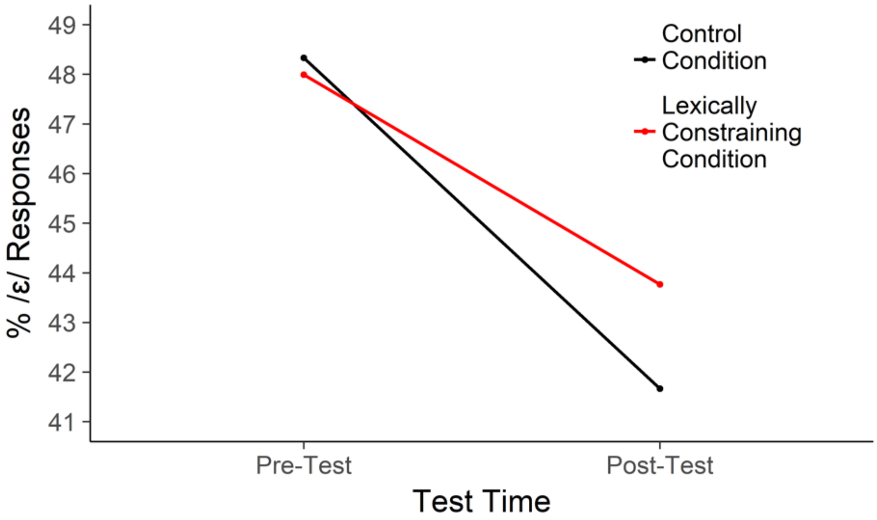

Figure 2. Mean percentage of /ɛ/ responses by condition and test time.

Given the unexpected finding that participants in both conditions shifted toward /ɪ/ responses in the post-test, we conducted post-hoc analyses to investigate whether the shift was due to mere exposure to the stimuli. The post-test /ɪ/ shift cannot be explained as a compensation for a pre-test bias in the opposite direction, as the percentage of items labeled as /ɪ/ was already 51.8% in the pre-test (significantly greater than half; $t(6479) = 2.96$, $p < 0.001$). Moreover, when trial number was added as a fixed effect to the model reported previously, it was significant in the opposite direction: in later trials within a test, responses tended more toward /ɛ/ ($\beta = 0.16$, $SE = 0.03$, $p < 0.001$). Therefore, the overall shift toward /ɪ/ in the post-test appears to result from either the time interval between the tests or from exposure to the speaker's voice during the interaction, rather than from repeated exposure to the vowel continuum.

5.4 Discussion

The phonetic categorization results include two main findings: first, listeners showed an overall tendency to shift toward /ɪ/ interpretations from the pre-test to the post-test, and second, this effect was attenuated for listeners in the lexically constraining condition: they made more /ɛ/ responses in the post-test than listeners in the control condition, specifically from the middle to the /ɪ/ half of the spectrum. Thus, listeners who were exposed to lexically constraining information show enough perceptual adaptation—learning that their interlocutor's /ɪ/ pronunciations actually represent /ɛ/ words—to partially counteract the larger /ɪ/-directional bias exhibited by both listener groups.

The relative subtlety of the observed lexically guided perceptual learning effect was in line with what we expected, given the continuous nature of vowel perception and the more cognitively demanding listening conditions of an L2 dialogue setting. However, the small effect size may also be due to a combination of the unexpected /ɪ/-shifting bias and several methodological factors. The overall slight bias toward /ɪ/ may reflect a well-documented asymmetry in vowel perception (Polka & Bohn, 2003): listeners can more easily discriminate a change from a more central vowel to a more peripheral vowel than vice versa since the more peripheral vowel serves as a perceptual anchor. Thus, when labeling sounds along the /ɛ/-/ɪ/ continuum, it is easier to hear that a given sound is more /ɪ/-like than the previous one, leading to a slight response bias in that direction. Why the preference for the /ɪ/ label increased between the pre-test and post-test might be because of the additional exposure to the speaker's voice during the Code Breaker interaction, in which only 16 /ɪ/ vowels and no /ɛ/ or /i/ vowels were heard. This manipulation in the set of vowels heard, or simply the additional information about the speaker's realization of other vowels, could have altered how listeners mapped her vowel space.

Several methodological aspects of the present study may also have contributed to the subtlety of the lexically guided adaptation. One is the relatively limited evidence presented for the vowel shift. After a long phonetic categorization pre-test with 216 items spanning the whole spectrum from /ɛ/ to /ɪ/, participants began the Code Breaker game with a well-founded expectation that their interlocutor would produce "e"-words with /ɛ/-like sounds. For participants to change their phonemic boundaries on the basis of just 16 critical accented words in the Code Breaker game thus requires a substantial amount of pre-test exposure to be unlearned. Strengthening the evidence for the vowel shift during the interactive game—whether by increasing the number of critical trials or by incorporating additional accented vowels to form a chain shift as in Maye et al.'s (2008) and Cooper and Bradlow's (2018) studies—would probably increase the lexically driven adaptation.

Another difference between our methodology and that of traditional lexically guided learning studies is that in our experiment, the learning was driven by lexical constraints built into the task itself, rather than from the listeners' mental lexicon. That is, due to the use of minimal word pairs as the

Code Breaker stimuli, all phonetic input listeners received during the game was compatible with real English words, even when those words were absent from the screen. Stronger perceptual learning may occur if listeners were to be exposed to accented words whose /ɪ/ pronunciations did not map onto real words (e.g., “best” pronounced /bɪst/), though this design would preclude the use of a control condition when using a cross-category sound shift rather than an ambiguous sound.

A strength of the present experimental design is that using a pre-test in addition to a post-test made it possible to assess perceptual adaptation within rather than only between listeners. Moreover, we have shown that employing pre-recorded speech in an interactive game is a fruitful method to study speech processing in a more naturalistic setting. In future research, it would be interesting to expand the present design to be able to directly compare the size of the effect for native and non-native listeners, for different types of sounds within the same speaker, or for listening conditions that differ in cognitive load.

5.5 Conclusions

The goal of this paper was to determine whether lexical information drives perceptual adaptation to a vowel shift for non-native listeners in an interactive context. To that end, participants played an interactive game containing lexical evidence that the interlocutor’s /ɪ/ pronunciation should be interpreted as /ɛ/, and their phonemic category boundaries were assessed with a phonetic categorization pre-test and post-test. Relative to the control condition, listeners in the lexically constraining condition were more likely to interpret /ɪ/-like sounds as /ɛ/ in the post-test, despite the fact that both listener groups were biased in the /ɪ/ direction. This shows that lexically guided perceptual adaptation can indeed occur for a vowel shift, from a relatively small amount of evidence, and within the cognitively demanding setting of an L2 task-based interaction, attesting to the robustness of this type of perceptual learning.

Chapter 6: How explicit instruction improves phonological awareness and perception of L2 sound contrasts in younger and older adults

Abstract:

Despite the importance of conscious awareness in second language acquisition theories, little is known about how L2 speech perception can be improved by explicit phonetic instruction. This study examined the relationship between phonological awareness and perception in Dutch younger and older adult L2 listeners, focusing on two English contrasts: a familiar contrast in an unfamiliar position (word-final /t/-/d/) and a harder, unfamiliar contrast (/æ/-/ɛ/). Awareness was assessed with a task in which written words belonging to homophone pairs and minimal pairs had to be judged as sounding the same or different. Perception was assessed with a two-alternative forced-choice identification task with auditorily presented minimal pairs. We investigated whether listeners' awareness and perception improved after a video-based explicit instruction that oriented their attention to one of these contrasts, and we tested whether including information about the phonetic cue of vowel duration increased learning. Awareness and perception of each contrast were shown to be moderately correlated at the study's outset. Furthermore, awareness and perception for each contrast generally improved more after the instruction drawing attention to that contrast. However, the effectiveness of explicit phonetic instruction was shown to vary depending on the combination of the contrast's difficulty, cue information, and listener age.

This chapter is based on the following:

Felker, E., Janse, E., Ernestus, M. & Broersma, M. (submitted). How explicit instruction improves phonological awareness and perception of L2 sound contrasts in younger and older adults.

6.1 Introduction

In second language (L2) speech perception, one challenge for late bilinguals is learning to distinguish sounds that are not contrastive in the native language (L1). Previous research has shown that intensive exposure to controlled stimuli using high variability phonetic training can improve adult listeners' ability to distinguish novel L2 sound contrasts (see Sakai & Moorman's meta-analysis, 2018). These perception training paradigms, which typically involve lengthy identification tasks with corrective feedback, are theorized to bring about changes in listeners' selective attention: over the course of training, listeners shift their attention to the acoustic-phonetic cues that are phonologically relevant for a given sound contrast (Francis et al., 2000; Francis & Nusbaum, 2002). Interestingly, there is much less research about the effectiveness of bringing relevant phonetic cues to listeners' awareness through explicit instruction, despite the importance of conscious awareness in theories of second language acquisition (e.g., Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994). The present study investigates how both awareness and perception can be improved by explicit instruction. Moreover, it expands upon previous research by including not only young adults but also older adult L2 listeners, who may show different learning effects than younger listeners due to age-related differences in cognitive processing abilities.

Explicit instruction can improve speech perception by orienting listeners' attention to what they need to learn. Several studies have shown that perception of unfamiliar phonemic contrasts can be improved by explicitly directing listeners' attention to sounds rather than semantics, or directing their attention to specific classes of sounds over others. Guion and Pederson (2007) exposed native English speakers to Hindi minimal word pairs based on Hindi stop consonant contrasts, along with the words' English translations; one participant group was told to attend to the words' sounds and the other to their meanings. For the most difficult contrast tested, the sound-attending group demonstrated greater perceptual discrimination improvement than the meaning-attending group. Similarly, Pederson and Guion-Anderson (2010) gave native English listeners identification training on Hindi words presented auditorily; listeners were instructed to attend to and identify either consonants or vowels. The consonant-attending group, but not the vowel-attending group, showed post-training improvement in consonant discrimination. The effectiveness of attention-directing has also been demonstrated for the learning of tonal contrasts. Chen and Pederson (2017) trained native Mandarin listeners on Quanzhou Southern Min words involving unfamiliar consonant and tonal contrasts. In the identification training, listeners were instructed to attend to and identify either the consonants or the tones. At post-test, the consonant-attending group had only improved in consonant discrimination, while the tone-attending group had only improved in tone discrimination. Taken together, these studies show that directing listeners' attention to the target sounds facilitates perceptual learning of unfamiliar sound contrasts. Whether such a simple intervention also works for L2 learners who have already had years of exposure to the non-native sound contrasts remains to be seen.

Some evidence suggests that perceptual learning of non-native sound contrasts benefits from focusing listeners' attention even more narrowly, to the level of specific phonetic cues. Hisagi and Strange (2011) showed that native English speakers were better at discriminating unfamiliar Japanese contrasts of vowel length, consonant length, and syllable length if they had first received written instructions explaining that duration was what made the words different. Similarly, Porretta and Tucker (2014) found that native English speakers were better at distinguishing unfamiliar Finnish consonants differing in length if they had first received basic written instructions pointing out the difference between short and long consonants. Drawing attention to specific phonetic cues can facilitate learning new sound categories even when the sounds to be learned differ in multiple dimensions, such as Mandarin tonal contrasts that differ in both pitch height and direction. Chandrasekaran et al. (2016) showed that a short written instruction telling native English listeners to focus on pitch direction, a dimension that they would normally underweight, improved their categorization of Mandarin tones more than instructions telling them to focus on pitch height or both direction and height. Thus, it appears that listeners can make use of explicit instruction about specific phonetic cues to improve their perception of unfamiliar non-native contrasts.

While the previous two studies provided no more than a few sentences of information about the phonetic cue, and tested listeners in an unfamiliar language, Kissling (2014) studied the effect of more intensive phonetic instruction with beginner, intermediate, and advanced learners of Spanish. Over multiple weeks, learners in the phonetic instruction group completed self-paced online modules about specific Spanish consonants that explained grapheme-phoneme correspondences, provided detailed articulatory phonetic instructions, and included sound identification exercises. The phonetic instruction group showed greater pre-to-post-test improvement in identification and discrimination of the target phones than a control group who completed modules with comparable sound exposure but no phonetics instruction. However, one potential confound was that the control group, unlike the phonetics instruction group, was never told which sounds were the target of the study. Thus, while exposure was controlled, the effect of the detailed phonetic information could not be separated from the effect of simply orienting learners' attention to the target sounds. In our study, we will examine whether explaining the phonetic cue of duration improves perception above and beyond orienting L2 listeners' attention to the critical sounds.

Second language speech perception has only rarely been studied in elderly listeners, whose speech processing differs from that of young adults in various ways due to age-related hearing loss, cognitive decline, and slowed temporal processing (see reviews of Gordon-Salant, 2005; Pichora-Fuller & Souza, 2003). For instance, Sommers (1997) found that older adults were less able than younger adults to ignore phonetically irrelevant stimulus dimensions in speech, implying a breakdown of selective attention. Moreover, older adults have shown less flexibility than younger adults in lexically guided perceptual category learning (Scharenborg & Janse, 2013). Older adults adapt as well as young adults to noise-vocoded speech, but only when given a less degraded signal to equate baseline accuracy between groups (e.g., Peelle & Wingfield, 2005; Neger et al., 2014). Furthermore, many studies suggest that while older listeners with normal hearing are quite capable of perceptual learning of speech, the benefits of training are varied and transfer of learning may be limited (see the review of Bieber & Gordon-Salant, in press). For instance, for time-compressed speech, older listeners have shown comparable perceptual learning to younger listeners when the age groups are equated for starting accuracy, but they did not transfer their learning as well as younger listeners to a different speech rate (Peelle & Wingfield, 2005). Overall, these studies show that older adults are capable of implicit perceptual learning but tend to show less selective attention, perceptual flexibility, and transfer of learning than younger adults.

To our knowledge, little research has compared older adults' and younger adults' perceptual learning of L2 speech. The existing crosslinguistic studies with elderly adults use languages that are unfamiliar to the listeners. For instance, older native Japanese speakers have shown improved perception after training on English syllable rhythm (Tajima et al., 2002) and English phonemic contrasts not present in Japanese (Kubo & Asahane-Yamada, 2006). The latter study, which included a direct comparison with younger adult learners, found a comparable learning effect in both age groups. More recently, Maddox et al.

(2013) and Ingvalson et al. (2017) investigated the ability of older adult native English speakers to learn to perceive Mandarin lexical tone categories based on identification training with corrective feedback. Maddox et al. (2013) reported that older adults performed worse overall and showed a lower learning rate than younger adults. Ingvalson et al.'s (2017) older adults also showed improved perception over the course of training, though their learning was not compared with younger adults. Together, these studies show that training can improve older listeners' perception of non-native sound contrasts, though older listeners may learn less effectively than younger listeners. It remains an open question how older listeners would respond to explicit phonetic instruction and whether such instruction might improve their perception of non-native contrasts in an L2 they are already proficient in, as opposed to an unfamiliar language.

6.1 The present study

The present study investigates how explicit instruction improves phonological awareness and perception of L2 sound contrasts by adult listeners. As described above, most previous studies about explicit phonetic instruction have focused on teaching non-native listeners about the sounds of a language that is completely unfamiliar to them, a scenario unlikely to ever occur outside the laboratory. We study the effect of instruction about L2 contrasts for adults who are already proficient in the L2 and whose phonemic categories may therefore be entrenched after years of language use. Furthermore, we vary the range of difficulty by testing the learning of two L2 sound contrasts that differ in their relation to the L1 sound system. Finally, we go beyond previous research by testing whether explicit instruction works not only for young adults but also for older adults, whose capacity for L2 sound learning and responsiveness to explicit phonetic instruction might be more limited due to various age-related differences in cognitive processing.

6.1.1 Research questions

We have three main research questions. First, we assess the relationship between L2 listeners' prior perceptual accuracy and their phonological awareness for each of the two contrasts, operationalized as the extent to which they know that minimal pairs based on the contrasts are meant to sound different (Research Question 1). Then, we investigate the effect of instruction that orients listeners' attention to one contrast or the other, testing whether the instruction improves listeners' phonological awareness (Research Question 2) and perceptual accuracy (Research Question 3) for the attended contrast. For Research Questions 2 and 3, we also test whether it is beneficial for the instruction to describe the phonetic cue of vowel duration, whether the duration information improves awareness or perception of the non-attended contrast, and whether learning differs between the two sound contrasts and listener age groups.

6.1.2 Study design

This study investigates the effect of explicit phonetic instruction on awareness and perception for Dutch younger and older adults, and it focuses on two English contrasts, word-final /t/-/d/ and /æ/-/ɛ/, which should pose differing degrees of difficulty for native Dutch listeners (see motivation below). The phonetic instruction was delivered through a short video in which a native English speaker described the contrast in question and drew attention to relevant minimal pairs. The focus on minimal pairs was inspired by the prominent role of minimal pairs in L2 teaching (Brown, 1995; Field, 2008) and in research about phonological awareness (e.g., Janssen et al., 2015; Krenca et al., 2020). With the instructional video, we aimed to test separately the effect of orienting listeners' attention to the critical sound contrast and orienting their attention to a specific phonetic cue. Therefore, participants were assigned to watch a video in one of four conditions: the video was either about the /t/-/d/ or the /æ/-/ɛ/ contrast, and it either did or did not explain how the phonetic cue of duration distinguished the sounds.

We chose to focus on Dutch listeners' perception of the English word-final /t/-/d/ contrast and the /æ/-/ɛ/ contrast, and duration as a phonetic cue to distinguish both contrasts, based on previous research. In English, a salient difference between word-final /t/ and /d/ is the preceding vowel's duration: English vowels typically shorten before voiceless consonants, like /t/, and lengthen before voiced consonants, like /d/ (House, 1961). For Dutch listeners, the English word-final /t/-/d/ contrast represents a familiar contrast in an unfamiliar position, as the Dutch devoicing of final obstruents allows /t/ but not /d/ in word-final position (e.g., Booij, 1999). Dutch listeners' perception of the final /t/-/d/ contrast is less accurate than that of English listeners in lexical processing (Broersma & Cutler, 2008). Moreover, experiments with phonetically manipulated stimuli have shown that while Dutch listeners are capable of exploiting vowel duration as a cue for word-final obstruent voicing contrasts, they do so to a lesser extent than native English listeners (Broersma, 2010).

The English /æ/-/ɛ/ contrast does not exist in Dutch at all and may thus be even more difficult for Dutch listeners to distinguish than /t/-/d/. Phonetically, /æ/ and /ɛ/ differ in both spectral frequency (Hillenbrand et al., 1995) and duration (Crystal & House, 1998), with /æ/ being longer than /ɛ/. In this phonetic space, the Dutch vowel system has only one vowel, transcribed as /ɛ/, whose phonetic realization falls between the English /æ/ and /ɛ/ (Collins & Mees, 1996). Dutch listeners have difficulty processing this contrast (Broersma, 2012), likely because they assimilate the English /æ/ and /ɛ/ to their native /ɛ/ category in accordance with the Perceptual Assimilation Model (Best & Tyler, 2007). Dutch listeners use duration to identify vowels in Dutch (Van der Feest & Swingley, 2011) and are capable of exploiting duration cues to categorize the English /æ/ and /ɛ/ in phonetically manipulated stimuli (Díaz et al., 2012).

To assess the effectiveness of the explicit phonetic instruction for both /t/-/d/ and /æ/-/ɛ/, we employed pre- and post-tests of two kinds: phonological awareness and perception. In the phonological awareness pre- and post-tests, a series of word pairs was presented orthographically. For each pair, participants had to indicate whether they thought the two words sounded the same or different. The more often they correctly classified the critical minimal pairs (e.g., *greet* and *greed*) as sounding different, the greater the phonological awareness for the contrast in question can be assumed. In the perception pre- and post-tests, each word from the critical minimal pairs was presented auditorily in the context of a two-alternative forced-choice listening task (e.g., hearing /bæg/ and having to label it as *bag* or *beg*), and we analyzed listeners' perceptual accuracy for each contrast.

6.1.3 Hypotheses

Research Question 1 concerns the relationship between phonological awareness of novel L2 sound contrasts and perceptual accuracy for those contrasts. The link between awareness and perceptual learning of L2 sound contrasts has not yet been empirically demonstrated, but awareness is theorized to play an important role in L2 acquisition in general (Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994). We therefore hypothesize that phonological awareness of each contrast (word-final /t/-/d/ or /æ/-/ɛ/) will correlate positively with perceptual accuracy for that contrast. We examine this awareness-perception relationship at pre-test in order to answer this question independently of the instructional intervention.

Research Question 2 is whether explicit phonetic instruction about a non-native contrast can increase L2 listeners' phonological awareness. Of course, awareness for both contrasts might increase from pre-test to post-test simply because the intervening perception task, which requires listeners to match each critical word they hear to one word label or another, implies that the critical minimal pair words are meant to sound different. However, we are mainly interested in whether any awareness gains are specifically due to the video instruction about the relevant sound. Crucially, we expect that /t/-/d/ awareness will increase more after watching a /t/-/d/ video than after an /æ/-/ɛ/ video, and /æ/-/ɛ/ awareness will increase more after watching an /æ/-/ɛ/ video than after a /t/-/d/ video. Whether providing information about the vowel duration cue will further increase awareness is uncertain. On the one hand, since the videos are already very explicit in stating that the sounds in question are distinctive, additional information about duration might be superfluous and therefore provide no added benefit. On the other hand, the explicit duration information might reinforce phonological awareness of either contrast by illustrating how the two sounds differ concretely.

Research Question 3 is whether explicit phonetic instruction about a non-native contrast can improve L2 listeners' perception. We predict that perception of /t/-/d/ will improve more after watching a /t/-/d/ video and that perception of /æ/-/ɛ/ will improve after watching an /æ/-/ɛ/ video. Unlike with phonological awareness, we do not expect perception of a contrast to improve after watching a video about the other contrast. Furthermore, we do predict that perception of a given contrast will improve more after a video describing the duration cue for that contrast than it will after a video without the duration cue information.

For both Research Questions 2 and 3, we expect that the younger adults will make more improvements in awareness and perception than the older adults, and we expect that more learning will occur for the easier /t/-/d/ contrast than for the more difficult /æ/-/ɛ/ contrast. We also expect that, if there is transfer of learning from one sound contrast to another, the perceived relevance of the phonetic cue could play a role. Specifically, the vowel duration cue in the /æ/-/ɛ/ video context may appear relevant only to vowel contrasts and therefore not be generalized to the /t/-/d/ contrast, whereas the vowel duration cue in the /t/-/d/ video context may be surprising enough to potentially suggest that the words in the vowel pairs should also sound different because of duration differences.

6.2 Methods

6.2.1 Participants

The participants were 124 monolingually-raised native Dutch speakers: 64 younger adults (72% female) aged 18 to 31 (mean = 22.3, *SD* = 2.9) years and 60 older adults with normal hearing[2] (62% female) aged 65 to 84 (mean = 69.9, *SD* = 4.1) years, who had started learning English on average at ages 10.7 (*SD* = 1.2, range: 8–14) and 12.3 (*SD* = 1.1, range: 10–16) years, respectively. The younger adults spent significantly more hours per week speaking and listening to English than the older adults: for speaking, mean 2.1 hours (*SD* = 4.1) vs. mean 0.6 hours (*SD* = 1.1), *t*(72.38) = 2.91, *p* = .005; and for listening, mean 12.4 hours (*SD* = 12.5) vs. mean 6 hours (*SD* = 6.5), *t*(96.85) = 3.60, *p* < .001). Additionally, the younger adults rated themselves higher than the older adults did on English proficiency across the skills of reading, writing, speaking, and listening (3.3 vs. 2.8 mean score across four scales from 0 “no ability” to 6 “perfect”; *t*(120.17) = 3.88, *p* < .001).

6.2.2 Materials

6.2.2.1 Perception test

The perception test consisted of 80 spoken English word tokens comprising 40 phonological minimal pairs: 20 word-final /t/-/d/ minimal pairs (e.g., *feet* and *feed*) and 20 /æ/-/ɛ/ minimal pairs (e.g., *bag* and *beg*), all listed in Appendix E (Table E1). The stimuli set only included words that we expected participants to know. To avoid potential interference in case the English words would activate similar-sounding words in the listeners’ L1, the stimuli set excluded words that sounded very similar to a Dutch word if their word-final /d/ was interpreted as /t/ or if their /æ/ was interpreted as /ɛ/, with three exceptions: *bead* (resembling Dutch *biet*), *bad* (resembling Dutch *bed*), and *had* (resembling Dutch *het*). The words were recorded in citation form by a female native speaker of

[2] Three additional older adults were tested but excluded from data analysis: one had a bilingual upbringing, one did not complete the experiment, and one demonstrated >35 dB of hearing loss (the threshold to qualify for hearing aids in the Netherlands).

Standard American English (the first author of this paper) in a sound-attenuating booth.

6.2.2.2 Phonological awareness test

The phonological awareness test comprised 100 English word pairs that were each one of two types: homophones or phonological minimal pairs. The critical items were the same 20 /t/-/d/ minimal pairs and 20 /æ/-/ɛ/ minimal pairs that were used in the perception test. By design, the critical pairs' type was uncertain: while they ought to be classified as minimal pairs, we expected many participants to misclassify them as homophones, at least in the pre-test. To keep the pair types relatively balanced from the participants' perspective, the filler items consisted of 30 homophone pairs (e.g., *son* and *sun*) and 30 minimal pairs involving various other vowel and consonant contrasts, in both onset and coda positions (e.g., *play* and *pray*), all listed in Appendix E (Table E2). The four pair types (/t/-/d/, /æ/-/ɛ/, and filler minimal pairs, plus the homophone pairs) were equivalent in their mean Zipf frequency in the SUBTLEX-US corpus (Brysbaert & New, 2009); $F(3, 196) = 0.22$, $p = .88$.

6.2.2.3 Phonetic instruction videos

We produced four different phonetic instruction videos: two about the word-final /t/-/d/ contrast and two about the /æ/-/ɛ/ contrast. For each contrast, one video explained the duration cue and the other did not. The four videos were recorded separately and featured the same native English speaker who had recorded the perception stimuli. Each critical contrast was illustrated with example words, including minimal pairs (e.g., *bit* and *bid*) and non-minimal pairs (e.g., *sit* and *did*) that did not occur in the pre- and post-tests and did not contain sounds from the other critical contrast. Whenever the speaker named a critical sound, a corresponding letter T, D, A, or E appeared briefly onscreen (Figure 1a), and when she pronounced an example word, it appeared briefly onscreen with the critical sound's letter darkened for emphasis (Figure 1b).

All videos were approximately two minutes long, and their scripts (see Appendix F) were as similar as possible in content, length, and structure. Crucially, all four videos contained the same number of instances of example words.

Each video comprised eight stages. First, the speaker introduced herself as a native speaker, named the critical sounds the video was about, and explained their contrastive role in English using two minimal pair examples. Second, using two non-minimal pair words, she stated the sounds' typical spellings. Third, she mentioned a subtle difference between the two sounds (aspiration for /t/-/d/ and vowel quality for /æ/-/ɛ/).

Fourth, the speaker either made a generic statement that the sounds were hard to distinguish for non-native listeners (no-cue videos) or she said that what would really help the viewer to hear the difference was to listen to how long the vowel was (duration-cue videos).

Fifth, the speaker said she was going to pronounce example words in an exaggerated way, and she pronounced two minimal pairs with either exaggerated aspiration or vowel quality (no-cue videos) or exaggerated short and long vowel length (duration-cue videos). In the duration-cue videos only, these exaggerated pronunciations were accompanied by hand gestures in which the palms began together and moved horizontally outward, either far beyond the body (for the longer /d/ and /æ/ words; Figure 1c) or shoulder-width apart (for the shorter /t/ and /ɛ/ words; Figure 1d).

Sixth, the speaker said she would pronounce the words more normally and invited the viewer to listen closely and try to hear the difference. In the duration cue videos only, she then described the duration cue explicitly (duration-cue videos) by stating that the vowel before /d/ sounded longer than the vowel before /t/ (/t/-/d/ duration-cue video) or by stating that /æ/ sounded longer than /ɛ/ (/æ/-/ɛ/ duration-cue video).

Seventh, she pronounced two minimal pairs, each repeated three times; this time, unlike before, the words did not appear onscreen until the third pronunciation to allow listeners to test their comprehension without the visual support.

Finally, the videos concluded by stating that the difference would become easier to hear with practice (no-cue videos) or by reiterating the duration difference for the relevant sound contrast (duration-cue videos; e.g., "just remember: the /æ/ sounds longer while the /ɛ/ sounds shorter").

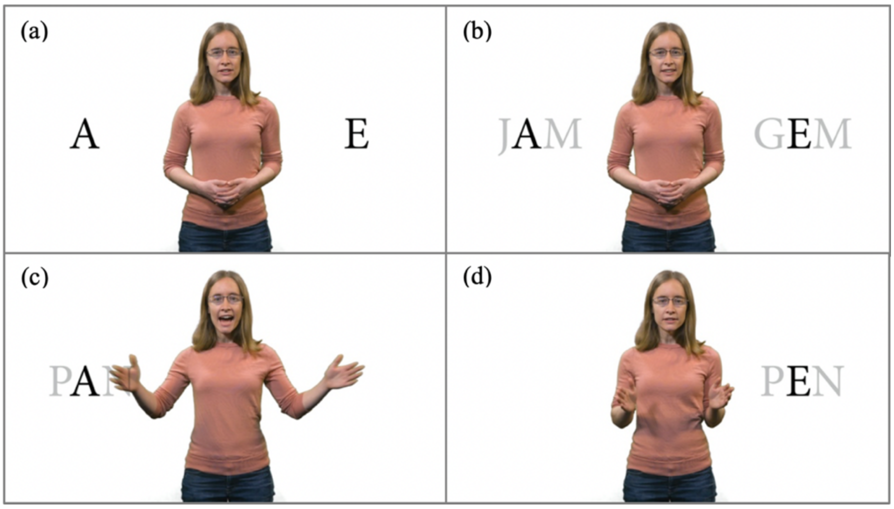

Figure 1. Stills from the /æ/-/ɛ/ duration-cue video.

6.2.3 Procedures

Participants were tested individually in a sound-attenuating booth. The older adults were first screened for hearing acuity with an Oscilla audiometer using an automated Hughson-Westlake procedure to obtain a pure-tone average threshold for each ear at 500, 1000, and 2000 Hz (air conduction only). All participants, older and younger adults alike, then completed the main tasks in the order shown in Figure 2, followed by a language background questionnaire.

To minimize interaction with the experimenter, thereby limiting extraneous speech exposure, participants received written onscreen English-language instructions for each task. They wore Sennheiser over-ear headphones for the perception tests and phonetic instruction video; volume was held constant across participants. Responses were made with a button box. All test trials were self-paced, and participants could take a short rest before playing the phonetic explanation video.

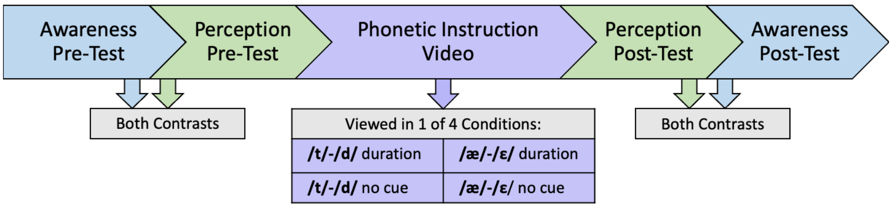

Figure 2. Order of the main tasks within each experimental session.

6.2.3.1 Perception pre-test and post-test

The two-alternative forced-choice perception pre-test and post-test each consisted of 80 trials containing the same 80 word tokens presented in a different randomized order for each participant and test time. In each trial, a minimal pair appeared with one word on each side of the screen, and the audio recording of the target word played automatically after a one-second delay. The instruction was to answer the question "Which word did you hear?" by pressing either the "left" or "right" labeled button. The left-right positioning for each critical sound in the /t/-/d/ and /æ/-/ɛ/ minimal pairs was counterbalanced across participants but held constant within each participant's session, thereby consistently associating each sound with one side of the screen.

6.2.3.2 Phonological awareness pre-test and post-test

The phonological awareness pre-test and post-test each consisted of the same 100 trials with a different randomized order for each participant and test time. In each trial, a word pair was displayed onscreen. The instruction was to answer the question "Do these words sound the same or different?" by pressing the button labeled "same" or "different." The word pairs' left-right positioning was counterbalanced across participants, and within each participant's session, words with the same critical sound always appeared on the same side of the screen.

6.2.3.3 Phonetic instruction video

Each participant watched one of the four phonetic instruction videos: the /t/-/d/ duration-cue video, the /t/-/d/ no-cue video, the /æ/-/ɛ/ duration-cue video, or the /æ/-/ɛ/ no-cue video. Videos were assigned to participants on a rotating basis, resulting in 16 younger adults and 14 to 16 older adults per video condition. During the video, participants had no other task than to pay attention.

After the video, onscreen text instructed participants to try to apply what they had learned in the post-tests.

6.3 Results

6.3.1 Relationship between awareness and perception (RQ1)

To assess the relationship between participants' phonological awareness and perception of the /t/-/d/ and /æ/-/ɛ/ contrasts, we analyzed their performance in the awareness and perception tasks at pre-test. For each task, we calculated each participant's overall accuracy for the /t/-/d/ and /æ/-/ɛ/ items separately. Then, we computed correlations between the awareness and perception of each contrast type, using Spearman's rank correlations since the scores were not normally distributed. Figure 3 presents the results visually. As predicted, the pre-test data revealed a significant positive correlation between phonological awareness and perceptual accuracy for both /t/-/d/ (ρ = 0.54, *p* < .001) and /æ/-/ɛ/ (ρ = 0.34, *p* < .001).[3]

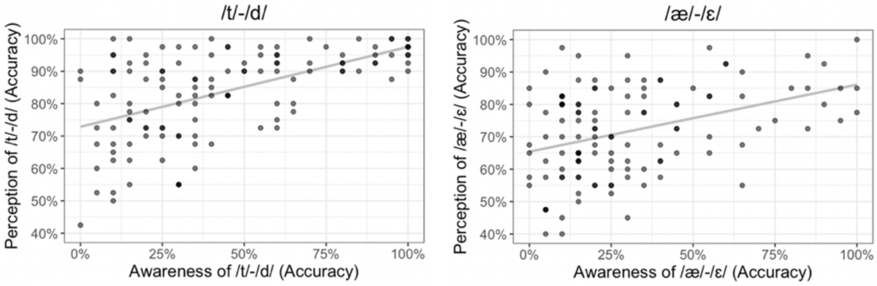

Figure 3. Pre-test correlations between phonological awareness and perceptual accuracy for /t/-/d/ (left) and /æ/-/ɛ/ (right).

[3] Similar correlation strengths were found when the awareness data was analyzed with d-prime scores rather than accuracy scores.

6.3.2 Effect of phonetic instruction on awareness (RQ2)

The phonological awareness results are presented graphically in Figure 4. To examine the effect of phonetic instruction on phonological awareness, we analyzed the accuracy of responses to the critical minimal pairs from the awareness pre-tests and post-tests using generalized linear mixed effects models with the logit link function from the *lme4* package in R (Bates et al., 2015). In these models, the binary dependent variable was accuracy (correct vs. incorrect). The random effects were item and participant (with random intercepts only, since random slopes prevented convergence). The fixed effects were age group (younger adults vs. older adults), contrast for the item in question (/t/-/d/ vs. /æ/-/ɛ/), test time (pre-test vs. post-test), video contrast (/t/-/d/ vs. /æ/-/ɛ/), video duration cue information (duration vs. no cue), and all possible interactions between these factors.

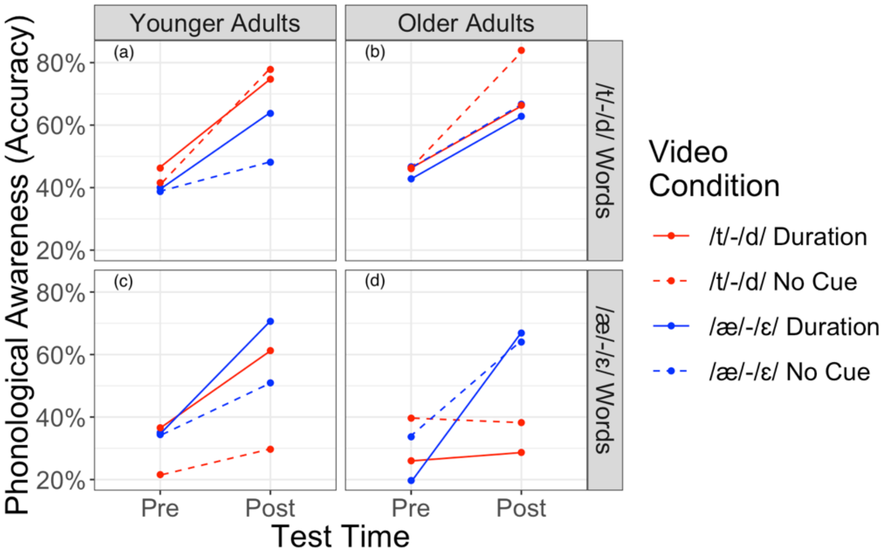

Figure 4. Changes in awareness from pre-test to post-test for the /t/-/d/ words (top row) and /æ/-/ɛ/ words (bottom row) for each combination of listener group, video contrast, and cue information.

The ANOVA table for the full statistical model is shown in Appendix G (Table G1). Since this model contained three significant four-way interactions and numerous significant lower-level interactions, we split the data by age group (which factored into all of the significant four-way interactions) and calculated separate models for the younger adults and older adults, as shown in Appendix G (Table G2). Again, these models contained many significant interaction effects, so to make their interpretation easier, we split the data for each group by item contrast (as this factored into all three of the significant three-way interactions for younger adults and into two of the three significant three-way interactions for older adults). The following subsections present the results in detail for each age group and item contrast.

6.3.2.1 Awareness in younger adults

For younger adults (see Figure 4a and 4c), the separate models for /t/-/d/ and /æ/-/ɛ/ awareness, with treatment coding, are presented in Table 1. For the /t/-/d/ items, younger adults showed a significant effect of test time: their accuracy was greater in the post-test than the pre-test, an effect which held for all four videos. As expected, they showed a significant interaction effect between test time and video contrast indicating that the post-test increase in accuracy was greater for the /t/-/d/ videos than for the /æ/-/ɛ/ videos. Additionally, a significant interaction between test time and cue information indicated that the pre-to-post-test improvement was greater for the duration-cue videos than the no-cue videos. The larger beta coefficient for the former interaction effect suggests that the matching video contrast was more beneficial than the presence of duration cue information.

For the /æ/-/ɛ/ items, young adults again showed a significant main effect of test time indicating that awareness increased from pre-test to post-test. The interaction between test time and video contrast, which would have shown the /æ/-/ɛ/ videos to lead to more improvement than the /t/-/d/ videos, did not reach significance (*p* = .08). However, there was a significant interaction between test time and cue information: the pre-to-post-test improvement was present for both types of videos but greater for the duration-cue videos than the no-cue videos.

Table 1

Models predicting awareness accuracy in younger adults (item contrasts separated)

Model for /t/-/d/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	-0.71	0.52	.17	[-1.73, 0.31]
Post-Test	0.66	0.21	.002*	[0.25, 1.07]
/t/-/d/ Video	0.20	0.71	.78	[-1.18, 1.59]
Duration Cue	0.21	0.70	.76	[-1.17, 1.59]
Post-Test • /t/-/d/ Video	1.62	0.31	<.001*	[1.02, 2.23]
Post-Test • Duration Cue	0.57	0.28	.04*	[0.02, 1.12]
/t/-/d/ Video • Duration Cue	-0.002	1.00	1.00	[-1.97, 1.96]
Post-Test • /t/-/d/ Video • Duration Cue	-0.54	0.44	.22	[-1.40, 0.32]
Random Effects	**Variance**			
Participant	3.63			
Item	0.40			

Model for /æ/-/ɛ/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	-1.70	0.47	<.001*	[-2.62, -0.78]
Post-Test	0.53	0.20	.01*	[0.13, 0.92]
/æ/-/ɛ/ Video	0.74	0.64	.25	[-0.52, 2.00]
Duration Cue	0.60	0.65	.35	[-0.67, 1.87]
Post-Test • /æ/-/ɛ/ Video	0.50	0.28	.08	[-0.05, 1.06]
Post-Test • Duration Cue	1.25	0.30	<.001*	[0.65, 1.84]
/æ/-/ɛ/ Video • Duration Cue	-0.49	0.91	.59	[-2.27, 1.30]
Post-Test • /æ/-/ɛ/ Video • Duration Cue	-0.24	0.42	.57	[-1.06, 0.58]
Random Effects	**Variance**			
Participant	2.91			
Item	0.25			

Note. SE = standard error, CI = confidence interval, * = significant.

6.3.2.2 Awareness in older adults

For older adults (see Figure 4b and 4d), the separate models for /t/-/d/ and /æ/-/ɛ/ awareness are presented in Table 2. For the /t/-/d/ items, the older listeners showed a significant simple effect of test time indicating higher accuracy in the post-test than in the pre-test. The significant two-way interaction between test time and video contrast, indicating greater pre-to-post-test improvement for the /t/-/d/ video condition, was only significant for the no-cue videos (mapped onto the intercept), as revealed by the signification three-way interaction between test time, video contrast, and cue information.

For the /æ/-/ɛ/ items, older listeners showed no significant simple effect of test time, but they did show a significant interaction between test time and video contrast, indicating that there was a significant pre-test to post-test improvement within the /æ/-/ɛ/ video condition but not within the /t/-/d/ video condition.

The fact that the three-way interaction between test time, video contrast, and cue information appears significant for /t/-/d/ but not for /æ/-/ɛ/ should be interpreted with caution, given that the four-way interaction including item contrast was not significant in the parent model in Appendix G (Table G2). Thus, it cannot be firmly concluded that duration cue information affects awareness gains differently for the /t/-/d/ items in the /t/-/d/ video condition than it does for the /æ/-/ɛ/ items in the /æ/-/ɛ/ video condition. Table 2 just suggests that the three-way interaction effect in the parent model indicating a negative effect of duration cue information is driven by the /t/-/d/ video condition for the /t/-/d/ items.

Table 2

Models predicting awareness accuracy in older adults (item contrasts separated)

Model for /t/-/d/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	-0.13	0.66	.83	[-1.42, 1.15]
Post-Test	1.41	0.23	<.001*	[0.97, 1.86]
/t/-/d/ Video	0.08	0.93	.93	[-1.73, 1.90]
Duration Cue	-0.29	0.89	.74	[-2.04, 1.46]
Post-Test • /t/-/d/ Video	1.63	0.37	<.001*	[0.90, 2.36]
Post-Test • Duration Cue	0.09	0.32	.77	[-0.54, 0.72]
/t/-/d/ Video • Duration Cue	0.11	1.29	.93	[-2.43, 2.64]
Post-Test • /t/-/d/ Video • Duration Cue	-1.41	0.51	.01*	[-2.40, -0.42]
Random Effects	**Variance**			
Participant	5.73			
Item	0.43			

Model for /æ/-/ɛ/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	-0.45	0.47	.34	[-1.38, 0.48]
Post-Test	-0.10	0.23	.64	[-0.55, 0.34]
/æ/-/ɛ/ Video	-0.47	0.63	.45	[-1.71, 0.76]
Duration Cue	-0.97	0.63	.13	[-2.21, 0.27]
Post-Test • /æ/-/ɛ/ Video	1.89	0.31	<.001*	[1.28, 2.49]
Post-Test • Duration Cue	0.29	0.31	.35	[-0.32, 0.90]
/æ/-/ɛ/ Video • Duration Cue	-0.03	0.88	.98	[-1.74, 1.70]
Post-Test • /æ/-/ɛ/ Video • Duration Cue	0.74	0.44	.09	[-0.12, 1.59]
Random Effects	**Variance**			
Participant	2.47			
Item	0.34			

Note. SE = standard error, CI = confidence interval, * = significant.

6.3.2.3 Summary of awareness results

For the young adult listeners, awareness of both the /t/-/d/ and /æ/-/ɛ/ contrasts increased from pre-test to post-test in all conditions, and it increased more after watching the duration-cue videos than after the no-cue videos. For the /t/-/d/ contrast, but not for the /æ/-/ɛ/ contrast, awareness also increased after watching any video specifically about that contrast.

For the elderly listeners, awareness of the /t/-/d/ contrast increased in all conditions but more from the /t/-/d/ videos than from the /æ/-/ɛ/ videos; moreover, it was specifically the /t/-/d/ no-cue video that raised /t/-/d/ awareness more than the other three videos. For the /æ/-/ɛ/ contrast, elderly listeners' awareness only improved at post-test from the /æ/-/ɛ/ videos.

6.3.3 Effect of phonetic instruction on perception (RQ3)

The perception results are presented graphically in Figure 5. To analyze the effect of phonetic instruction on perceptual accuracy, we computed generalized linear mixed effects models using the same model structures as described above for the awareness data analysis. The ANOVA table for the full statistical model is presented in Appendix H (Table H1). As the five-way interaction between all of the factors was significant, we again split the data by age group. The ANOVA tables for each age group's separate model are shown in Appendix H (Table H2). For both age groups, the item contrast factored into the highest-level significant interaction effect (a three-way interaction for younger adults and a four-way interaction for older adults), so we split the data further by item contrast. The following subsections describe in detail the results for each age group and item contrast.

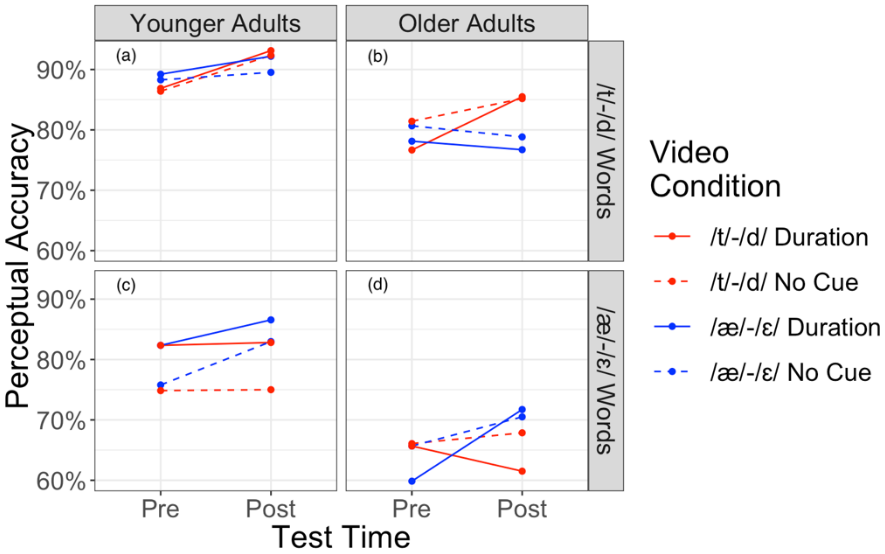

Figure 5. Changes in perceptual accuracy from pre-test to post-test for the /t/-/d/ words (top row) and /æ/-/ɛ/ words (bottom row) for each combination of age group, video contrast, and cue information.

6.3.3.1 Perception in younger adults

For younger adults (see Figure 5a and 5c), the separate models for perceptual accuracy for /t/-/d/ and /æ/-/ɛ/ items are presented in Table 3. The only significant effect within the /t/-/d/ model is a significant two-way interaction between test time and video contrast indicating that listeners in the /t/-/d/ video condition, but not those in the /æ/-/ɛ/ video condition, improved in t/-/d/ accuracy from the pre-test to the post-test. Similarly, within the /æ/-/ɛ/ model, the only significant effect was the interaction between test time and video contrast, indicating that listeners in the /æ/-/ɛ/ video condition, but not those in the /t/-/d/ condition, improved in /æ/-/ɛ/ accuracy from the pre-test to the post-test. Neither model showed any significant effects involving cue information.

Table 3

Models predicting perception accuracy in younger adults (item contrasts separated)

Model for /t/-/d/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	2.72	0.34	<.001*	[2.07, 3.39]
Post-Test	0.16	0.20	.41	[-0.22, 0.55]
/t/-/d/ Video	-0.30	0.41	.47	[-1.11, 0.51]
Duration Cue	-0.05	0.42	.91	[-0.86, 0.77]
Post-Test • /t/-/d/ Video	0.60	0.28	.03*	[0.05, 1.15]
Post-Test • Duration Cue	0.25	0.28	.39	[-0.31, 0.80]
/t/-/d/ Video • Duration Cue	0.26	0.59	.66	[-0.89, 1.41]
Post-Test • /t/-/d/ Video • Duration Cue	-0.12	0.41	.77	[-0.91, 0.68]
Random Effects	Variance			
Participant	1.05			
Item	1.01			

Model for /æ/-/ɛ/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	1.27	0.23	<.001*	[0.83, 1.72]
Post-Test	0.01	0.14	.94	[-0.26, 0.28]
/æ/-/ɛ/ Video	0.08	0.29	.77	[-0.48, 0.64]
Duration Cue	0.56	0.29	.05	[-0.01, 1.13]
Post-Test • /æ/-/ɛ/ Video	0.49	.20	.01*	[0.10, 0.89]
Post-Test • Duration Cue	0.03	0.21	.90	[-0.38, 0.44]
/æ/-/ɛ/ Video • Duration Cue	-0.14	0.41	.73	[-0.95, 0.66]
Post-Test • /æ/-/ɛ/ Video • Duration Cue	-0.17	0.30	.56	[-0.77, 0.42]
Random Effects	Variance			
Participant	0.50			
Item	0.41			

Note. SE = standard error, CI = confidence interval, * = significant.

6.3.3.2 Perception in older adults

For older adults (see Figure 5b and 5d), the separate models for perceptual accuracy for /t/-/d/ and /æ/-/ɛ/ items are presented in Table 4. Within the /t/-/d/ model, the only significant effect was the two-way interaction between test time and video contrast, indicating that listeners in the /t/-/d/ video condition, but not the /æ/-/ɛ/ video condition, improved in /t/-/d/ accuracy from pre-test to post-test. Within the /æ/-/ɛ/ model, the only significant interaction was the three-way interaction between test time, video contrast, and duration cue, which shows that only listeners who watched the /æ/-/ɛ/ duration-cue video improved in /æ/-/ɛ/ accuracy from pre-test to post-test.

6.3.3.3 Summary of perception results

The perception models show that both the /t/-/d/ and /æ/-/ɛ/ videos improved young adult listeners' perception of the featured contrast, regardless of whether the duration cue was mentioned, whereas their perception did not improve at post-test for the contrast not featured in the video. The elderly adult listeners performed similarly to the young adults for the /t/-/d/ contrast, demonstrating improved perception at post-test after watching either /t/-/d/ video. However, for the /æ/-/ɛ/ contrast, their perceptual learning was more limited: elderly listeners only improved at post-test after the /æ/-/ɛ/ duration-cue video.

Table 4

Models predicting perception accuracy in older adults (item contrasts separated)

Model for /t/-/d/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	1.91	0.37	<.001*	[1.17, 2.64]
Post-Test	-0.14	0.16	.37	[-0.45, 0.17]
/t/-/d/ Video	0.25	0.50	.62	[-0.73, 1.24]
Duration Cue	-0.16	0.48	.74	[-0.10, 0.79]
Post-Test • /t/-/d/ Video	0.49	.24	.04*	[-0.02, 0.96]
Post-Test • Duration Cue	0.04	0.22	.86	[-0.39, 0.47]
/t/-/d/ Video • Duration Cue	-0.23	0.70	.74	[-1.60, 1.13]
Post-Test • /t/-/d/ Video • Duration Cue	0.39	0.33	.24	[-0.26, 1.04]
Random Effects	**Variance**			
Participant	1.58			
Item	0.77			

Model for /æ/-/ɛ/ Items

Fixed Effects	B	*SE*	*p*-value	95% CI
(Intercept)	0.78	0.20	<.001*	[0.39, 1.18]
Post-Test	0.09	0.14	.49	[-0.17, 0.36]
/æ/-/ɛ/ Video	-0.05	0.24	.84	[-0.53, 0.43]
Duration Cue	-0.03	0.24	.91	[-0.51, 0.45]
Post-Test • /æ/-/ɛ/ Video	0.16	0.19	.41	[-0.21, 0.53]
Post-Test • Duration Cue	-0.30	0.19	.11	[-0.67, 0.07]
/æ/-/ɛ/ Video • Duration Cue	-0.25	0.34	.45	[-0.92, 0.41]
Post-Test • /æ/-/ɛ/ Video • Duration Cue	0.65	0.26	.01*	[0.14, 1.16]
Random Effects	**Variance**			
Participant	0.30			
Item	0.41			

Note. SE = standard error, CI = confidence interval, * = significant.

6.4 Discussion

The purpose of this study was to test the relationship between phonological awareness and perception of non-native contrasts and to investigate whether explicit phonetic instruction increases awareness and improves perception in both younger and older adult L2 listeners. To this end, we analyzed Dutch listeners' awareness and perception of two difficult English contrasts (word-final /t/-/d/ and /æ/-/ɛ/) both before and after they watched a short video that used minimal pairs to explain one of the contrasts, either /t/-/d/ or /æ/-/ɛ/, and that either did or did not explain the phonetic cue of duration.

First, we assessed the relationship between phonological awareness and perception by determining their correlation for each of the L2 sound contrasts at the outset of the experiment. In our study, awareness was operationalized as the proportion of minimal pairs with the contrast (presented orthographically) for which the two words were correctly judged as sounding different. The results showed that, as hypothesized, there were positive correlations between /t/-/d/ awareness and /t/-/d/ perceptual accuracy and between /æ/-/ɛ/ awareness and /æ/-/ɛ/ perceptual accuracy. While second language acquisition research has theorized that L2 learning is closely linked to being consciously aware of aspects of L2 form (Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994), to our knowledge this is the first time that the link between awareness and perception of specific L2 sound contrasts has been established. This correlation might arise because being able to perceive the difference between two L2 sounds makes people more likely to label them as different, and conversely, having awareness that two L2 sounds are meant to be different may be a crucial step on the path of learning to perceive that difference.

Second, we examined whether the explicit phonetic instruction increased younger and older adults' phonological awareness of the two contrasts. The main question was whether phonological awareness would increase more for the sound contrast that was featured in the instructional video. This beneficial effect of attention orienting was indeed borne out for the /t/-/d/ contrast in younger adults and for the /t/-/d/ and /æ/-/ɛ/ contrast in older adults: in these cases, awareness for the given contrast improved from pre-test to post-test more among those who had watched a video about that contrast than among those who had watched a video about the other contrast. This phonological awareness finding mirrors the perception-related findings of Pederson and Guion-Anderson (2010) and Chen and Pederson (2017), who demonstrated that directing attention to a particular contrast improves perception of the attended contrast more than the non-attended contrast. The fact that our younger adults showed only a marginally (p = 0.08) significant effect of video contrast for /æ/-/ɛ/ seems to arise from the strength of the duration-cue transfer effect: specifically, the fact that their /æ/-/ɛ/ awareness, which improved to some degree in all four video conditions, also improved remarkably from the /t/-/d/ duration-cue video. Thus, hearing about vowel duration differences, even in the context of a consonant contrast, may have been enough to trigger awareness that words with the critical vowels also ought to sound different.

When it comes to the direct effect of the duration cue on awareness-raising, the effects differed by age group. Young adults gained more awareness about both /t/-/d/ and /æ/-/ɛ/ from the duration-cue videos than from the no-cue videos. As these duration-cue effects for both contrasts did not interact significantly with video contrast, it appears that learning about the vowel duration cue increased awareness for both contrasts regardless of the context in which the cue was presented. The placement of the perception post-test between the video instruction and awareness post-test could have supported this generalization of learning by making the younger adults more likely to notice vowel length in both the /t/-/d/ and /æ/-/ɛ/ words they heard, which subsequently made them more likely to classify them as sounding different. The older adults, in contrast to the younger adults, did not benefit from the duration cue for either contrast. In fact, while the effect of the cue information on /æ/-/ɛ/ is unclear, the older adults' awareness of the /t/-/d/ contrast seemed to improve significantly more after watching the /t/-/d/ no-cue video than after the /t/-/d/ duration-cue video. Thus, not only did the vowel duration information provide no added benefit, it may have even been confusing or distracting for older adults, at least when presented in the consonant context where it may have had less perceived relevance. A practical implication of these results is that we do not endorse a one-size-fits-all approach to explicit phonetic instruction, as it seems that younger and older adult L2 listeners do not necessarily make the same use of additional phonetic cue information.

In addition to the aforementioned age-group differences in how phonological awareness is affected by the duration cue, there was one more age effect in the awareness gains: while the younger adults' awareness of both contrasts improved from pre-test to post-test in all four video conditions, the older adults' awareness gains were more limited. Specifically, the older adults' /æ/-/ɛ/ awareness did not improve at all after watching a /t/-/d/ video. Thus, while the older adults did gain awareness from the /æ/-/ɛ/ attention-orienting instruction as expected, they did not gain /æ/-/ɛ/ awareness simply through completing the intervening perception tests nor through transferring vowel length information from the /t/-d/ duration-cue video. This matches our expectation that awareness for the more difficult of the two contrasts, especially in the absence of an explicit instruction orienting attention to that contrast, would be less likely to increase for the age group that, for perceptual learning, tends to show limited transfer of learning to untrained stimuli (Bieber & Gordon-Salant, in press; Peelle & Wingfield, 2005).

Finally, we examined whether explicit phonetic instruction improved younger and older adults' perception of the two contrasts. The most important question was whether perception of the L2 contrast would improve after receiving explicit phonetic instruction about that contrast, which would attest to a positive effect of attention orienting on perception. This effect was clearly borne out for younger adults: they improved in /t/-/d/ perception after receiving /t/-/d/ instruction, but not after /æ/-/ɛ/ instruction, and they improved in /æ/-/ɛ/ perception after receiving /æ/-/ɛ/ instruction, but not after /t/-/d/ instruction. These results align with previous studies showing that perception of sounds in an unfamiliar language improves after being instructed to focus on those sounds specifically during training (Pederson & Guion-Anderson, 2010; Chen & Pederson, 2017). We have shown that this effect also holds for highly proficient L2 listeners. Our older adults' /t/-/d/ perception also improved after /t/-/d/ instruction but not after /æ/-/ɛ/ instruction. In contrast, their /æ/-/ɛ/ perception did not improve more from /æ/-/ɛ/ instruction compared to /t/-/d/ instruction overall. However, a significant interaction between test time, video contrast, and cue information showed that their /æ/-/ɛ/ perception did improve from the /æ/-/ɛ/ duration-cue instruction, unlike in the other three conditions. This more limited learning for the /æ/-/ɛ/ contrast in older listeners aligns with our expectation that the more challenging contrast would be less prone to improvement in the older age group.

We had hypothesized that providing information about the vowel duration cue for a given sound contrast would improve perception of that contrast, for both listener groups. However, the only significant effect of duration information on perceptual improvement was the aforementioned benefit of the /æ/-/ɛ/ duration-cue instruction over the /æ/-/ɛ/ no-cue instruction for older listeners. Thus, the duration cue was only helpful for perception in the most difficult listening condition. Moreover, there were no transfer effects showing that duration cue information from one contrast improved perception of another contrast. Interestingly, the previous studies showing perceptual learning in response to explicit phonetic instruction about duration (Hisagi & Strange, 2011; Porretta & Tucker, 2014) involved relatively challenging listening conditions, as they tested listeners in an unfamiliar non-native language on sounds that *only* differed in duration. In contrast, duration is just one of multiple cues that distinguishes the contrasts in this study, and our listeners already had a great deal of exposure to and proficiency in the language containing the sound contrasts. While Chandrasekaran et al. (2016) demonstrated a benefit of instructing non-native listeners about a phonetic cue that was absent in their native language, the vowel duration cue in our study is very prominent in our listeners' native language (Booij, 1999). All of this suggests that the perceptual benefit of instruction about a specific phonetic cue, above and beyond mere attention-orienting to the contrast, may be most likely to arise when the L2 contrast and/or phonetic cue are relatively difficult in light of the listeners' native language.

As mentioned above, the only age-related difference in perceptual learning was that the older adults showed more limited improvement for /æ/-/ɛ/ than the young adults by failing to improve from the /æ/-/ɛ/ no-cue instruction. This aligns with our expectation that less perceptual learning would take place for the more difficult contrast in the older listener group. While our older adults showed more restricted /æ/-/ɛ/ perceptual learning than younger adults, both age groups showed the same pattern of results for /t/-/d/ perceptual learning, consistent with Kubo and Asahane-Yamada's (2005) findings of equivalent learning between older and younger adults given perceptual training on L2 phonetic contrasts. Interestingly, despite previous research attesting to older listeners' breakdown of selective attention for phonetically relevant stimulus dimensions in speech (Sommers, 1997), the older listeners in our study were the only ones who benefited from the duration-cue phonetic instruction over the no-cue instruction for /æ/-/ɛ/. Thus, despite the negative effect of the vowel duration information on older adults' *awareness*, at least for the /t/-/d/ contrast, such information was apparently helpful for their *perceptual* learning of the /æ/-/ɛ/ contrast. This suggests that when a phonetic cue has high perceived relevance, as vowel duration does for a vowel contrast, older listeners are quite capable of using it to improve their perception of an L2 contrast.

Overall, this study's explicit phonetic instruction combined multiple components that each potentially contributed to the awareness gains and perceptual category learning: the presentation of minimal word pairs involving the critical L2 sounds, the description of the sounds' contrastive role, the phonetic cue information, the listening practice with exaggerated pronunciations, and even the exposure to the native speaker's voice. This study varied the presence of the duration cue information, following up on previous research about the benefits of instruction about non-native phonetic cues (Chandrasekaran et al., 2016; Hisagi & Strange, 2011; Porretta & Tucker, 2014). Further work is needed to determine which of the instruction's other elements also impacts phonetic learning. Moreover, future research could determine the extent to which the learning effects observed here will generalize to listeners' perception of other speakers and how long the awareness and perceptual gains will persist.

In conclusion, we have shown that a brief explicit phonetic instruction can improve phonological awareness and perception of L2 sound contrasts in younger and older adult listeners. In doing so, we tested two L2 contrasts that varied in difficulty and investigated the effect of including information about the phonetic cue of vowel duration in the explicit instruction. First, we established the correlation between awareness and perception of specific L2 contrasts at the outset of the experiment. Second, we demonstrated that phonological awareness generally increased more for the contrast that was featured in the instruction, thereby showing that attention-orienting enhances awareness. Moreover, while younger adults generalized the phonetic cue information to also increase their awareness of a non-attended contrast, older adults showed few transfer effects in awareness. Finally, we showed that for younger adults, explicit phonetic instruction for a given contrast improved perception of that contrast, regardless of the inclusion of the duration cue. For older adults, instruction improved perception of the easier contrast, regardless of cue information, whereas instruction improved perception of the more difficult contrast only when the duration cue was provided. Altogether, these findings shed new light on the conditions under which explicit instruction can orient L2 listeners' attention and improve their speech perception, revealing several important interactions between the specific L2 contrasts in question, the phonetic content of the instruction, and the listeners' age.

Chapter 7: General Discussion

This dissertation was about the learning of L2 speech perception in natural contexts, such as conversational interaction or a simple phonetic instruction. Specifically, it focused on how L2 listeners can improve their word recognition ability by perceptually adapting to an unfamiliar accent and by learning to better distinguish a non-native sound contrast. The aim of the dissertation was twofold: first, to present and evaluate two ecologically valid methods for studying speech processing, and second, to investigate three learning mechanisms for L2 speech perception: implicit lexical guidance, interactional corrective feedback, and explicit phonetic instruction.

This chapter first discusses the methodological contributions (Section 7.1) and theoretical findings (Section 7.2) of the preceding chapters, next suggests practical implications for the teaching of L2 speech perception (Section 7.3), then proposes directions for future research (Section 7.4), and finally summarizes the conclusions of the dissertation (Section 7.5).

7.1 Methodological contributions to studying more natural language processing

In recent years, researchers in psycholinguistics and related fields have called for studying language processing with more ecologically valid methods by using more natural types of speech stimuli, such as connected speech from casual conversation, and by focusing on more natural language use settings, such as interactive dialogue (e.g., Tanenhaus & Brown-Schmidt, 2008; Tucker & Ernestus 2016; Willems 2017). Dictation tasks based on connected speech are often employed in the context of L2 learning (e.g., Buck, 2001; Oller & Streiff, 1975; Savignon, 1982; Stansfield, 1985), but the traditional scoring methods based on lexical error rate alone do not fully capture the phonetic and semantic aspects of the input that listeners can recover. Moreover, while conversational interaction is thought to be a crucial locus of L2 acquisition (Ellis, 1999, 2003; Long, 1980), the learning of L2 phonology has almost never been studied in interactive dialogue due to challenge of sufficiently controlling the phonetic input. In this dissertation, Chapter 2 demonstrated how the dictation task could become a more valuable linguistics research tool by employing more reliable and informative scoring measures. Furthermore, Chapter 3 presented a novel experimental paradigm for studying speech processing in dialogue, combining the convincing illusion of a live conversation with the phonetic control afforded by using pre-recorded speech.

7.1.1 Studying the perception of conversational speech with the dictation task

Chapter 2 described and evaluated four different measures that can be used to score the transcriptions made in a dictation task, in this case one in which L1 and L2 listeners transcribed short stretches of conversational speech. The following four scoring measures were evaluated: lexical error rate, orthographic edit distance, phonological edit distance, and semantic error rate. First, the discriminative validity of the measures was supported by results showing that the L1 listeners significantly outperformed the L2 listeners across all four measures. Such a difference was expected given that L2 listening is generally more difficult than L1 listening (Cutler, 2012) and given prior research showing that reduced pronunciations, which were prevalent in this study's casual speech excerpts, are especially difficult for L2 listeners (e.g., Brand & Ernestus, 2018; Ernestus et al., 2017). Second, in addition to differentiating the two listener groups, the measures differed from each other when it came to quantifying listeners' performance: overall, listeners performed worst in terms of lexical error rate, better in terms of semantic error rate and orthographic edit distance, and best in terms of phonological edit distance. This suggests that lexical error rate, the scoring measure traditionally used in foreign language teaching contexts (Buck, 2001), may underestimate listeners' competence when it comes to their ability to recover the semantics and sounds of speech. Third, the criterion validity of the four measures was supported by the finding that the measures were significantly correlated with listeners' self-rated L2 proficiency, weekly hours of English listening (but not speaking), and L2 proficiency as measured by LexTALE (Lemhöfer & Broersma, 2012). Finally, the four measures were highly correlated with each other, with the lowest intercorrelation being between the semantic and phonological measures, as expected.

Which measure or combination of measures to use for speech perception research should be based on both practical and theoretical considerations. Chapter 2 demonstrated how lexical error rate, orthographic edit distance, and phonological edit distance can be calculated programmatically, which is advantageous because the potential subjectivity and human error in scoring transcriptions is greatly reduced. Lexical error rate may be the most relevant measure for assessing listeners' word recognition or segmentation abilities, though its sensitivity to spelling accuracy and binary nature (a transcribed word matches a target word either exactly or not at all) are drawbacks. The two gradient measures, orthographic and phonological edit distance, are more informative about the accuracy of listeners' phoneme perception. Of these measures, the phonological edit distance is the most complicated calculation, requiring a word-to-phoneme dictionary and grapheme-to-phoneme engine. While such resources are readily available for the English language, they may be harder to obtain for other languages that researchers are interested in. The orthographic edit distance, in contrast, is simple to calculate and still improves upon the lexical error rate by giving more credit to misperceived words that are spelled similarly to (and thus, to some extent, sound similar to) the target words. While many studies on the perception of continuous, accented speech have simply analyzed lexical error rate in listeners' transcriptions (e.g., Bent & Bradlow, 2003; Bradlow & Bent, 2008) or in their oral repetition of target phrases (e.g., Mitterer & McQueen, 2009; Pinet, Iverson & Huckvale, 2011), applying the orthographic edit distance to written transcriptions could enrich these types of data analyses by providing additional information about perceptual accuracy at the sound level. This approach has been applied in recent research that I supervised about the effects of subtitles on the perceptual learning of accented speech in an L2 (Van Gasteren, 2019, following up on research by Mitterer & McQueen, 2009).

Lastly, the one non-automatically calculated measure, semantic error rate, has the practical limitation of relying on the rater's own judgment as to whether key conceptual elements in the target phrase are present in the listeners' transcriptions. While Chapter 2 proposed some specific criteria to facilitate these determinations, the semantic error rate remains the most subjective and time-consuming measure to calculate. Nevertheless, it is a theoretically relevant measure for assessing listening comprehension, particularly in applied linguistics research contexts. The semantic error rate measure can capture the extent to which listeners grasp the meaning or gist of an utterance, even when they miss out on minor details of morphology or syntax. The desire to quantify speech perception ability in listeners who have not yet mastered the syntactical and morphological complexities of a language may arise frequently in L2 acquisition research. In particular, lower-proficiency L2 listeners have been shown to have trouble attending to form and meaning simultaneously and tend to process input primarily for meaning until their comprehension is sufficiently automatized (VanPatten, 1990; Wong, 2001). Furthermore, as L2 listeners use both bottom-up and top-down processing strategies for listening (Field, 2008; O'Malley et al., 1989; Vandergrift, 2007), the semantic error rate could effectively complement the phonological edit distance to enable researchers to measure listening performance based on both types of strategies.

7.1.2 Studying language processing in conversation with the ventriloquist paradigm

Chapter 3 presented a newly developed experimental method called the ventriloquist paradigm that enables the study of speech processing in conversational interaction with full control over phonetic exposure. In this paradigm, participants interact face-to-face with a confederate who communicates by playing pre-recorded speech recordings, including task-relevant phrases and flexible utterances belonging to pre-determined categories, into participants' headphones, all while briefly hiding her face behind her computer screen whenever she is "speaking." First, the validity of the paradigm was established by analyses covering over one hundred experiment sessions and two different confederates with two different pre-recorded speakers. Overall, four out of five participants reported no suspicion that their interlocutor's speech was pre-recorded and were thus likely convinced they were having a genuine conversation. Second, results showed that in an alternative setup where the confederate and participant sat in separate testing booths, rather than face-to-face, only one in three participants believed they were interacting with a real person whereas the remainder believed they were interacting with a computer or robot. Thus, the face-to-face setup was apparently crucial for upholding the illusion that the pre-recorded speech was live. Third, the ventriloquist paradigm was shown to lead to more engaging conversation, as measured by participants' speaking behavior, than a comparable setup in which participants were told upfront that they were only interacting with a computer.

When using the ventriloquist paradigm, making the interaction natural and believable requires careful preparation but is crucial in order for the interactive setting to have ecological validity. Insights from social psychology, such as Kuhlen and Brennan's (2013) warning that scripted confederates should not violate participants' pragmatic expectations, can be very informative when designing the tasks and scripted speech for ventriloquist paradigm experiments. Several ways to increase the naturalness and believability of the ventriloquist interaction beyond the experiments reported in Chapter 3 were implemented in a more recent ventriloquist paradigm experiment that I supervised, which convinced nine in ten participants that the interaction was genuine (Ye, 2020). In this implementation, the scripting of the pre-recorded speech took inspiration from recordings of speech made by real participants during similar experiments in our lab. In addition, this time we put the trials of the cooperative task in a fixed order, rather than a random one, which made the pre-recorded speech more realistic in three ways: the speaker could use anaphoric expressions, her sentence patterns and speech rate could change over time as she ostensibly familiarized herself with the task, and her intonation could flow more naturally between consecutive trials. Finally, the task and script were revised to increase interactivity by making the confederate sometimes provide underspecified information, predictably inducing the participant to ask clarification questions that would in turn elicit pre-determined follow-up utterances (e.g., "Now we need a star."—"What color?"—"An orange one."). Increasing the task's interactivity could plausibly increase the believability of the ventriloquist's speech being produced in real time, given the finding in Chapter 3 that participants were *less* likely to report suspicions of pre-recorded speech in sessions where the ventriloquist played *more* pre-recorded utterances. While the ventriloquist paradigm might never be able to create a convincing illusion for one hundred percent of participants, the successful adaptations made by Ye (2020) are a promising sign that the paradigm's naturalness and believability can be continually improved.

Chapter 3's comparison between the ventriloquist paradigm and the computer-interlocutor control setup raises interesting questions about the theoretical value of each experiment type. The computer-interlocutor setup is a useful method in its own right as a "Wizard of Oz" experiment (Fraser & Gilbert, 1991; Riek, 2012), a technique for studying human-computer interaction. While this setup is somewhat easier to implement than the ventriloquist paradigm, it should not be considered a substitute for studying human speech processing in face-to-face interaction. A growing body of recent research is showing that changes in speech processing, including various aspects of phonetic alignment (e.g., Burnham et al., 2010; Cohn et al., 2019; Raveh et al., 2019; Zellou & Cohn, 2020) and perceptual adaptation (Segedin et al., 2019), differ depending on whether people believe they are interacting with a human or a computer. While some degree of linguistic alignment may arise from automatic priming processes, other underlying mechanisms for linguistic alignment, such as audience design and social affective considerations, may be mediated by beliefs about one's interlocutor (Branigan et al., 2010). Therefore, the ventriloquist paradigm is more suitable for answering research questions about how speech processing is affected by social, affective, and communication-related aspects of human-human conversational interaction. In Section 7.4, specific suggestions for further research employing this paradigm will be presented.

7.2 Learning mechanisms for L2 speech perception

This dissertation investigated three different learning mechanisms for L2 speech perception: implicit lexical guidance, interactional corrective feedback, and explicit phonetic instruction. Lexical guidance has primarily been studied in non-interactive listening contexts (see reviews of Samuel & Kraljic, 2009 and Baese-Berk, 2018), and similarly, corrective feedback for speech perception has primarily been studied with intensive computer-based training programs (e.g., Bradlow et al., 1999; Iverson et al., 2005; Lee & Lyster, 2016b; Wang & Munrow, 2004). Chapters 4 and 5 investigated the effectiveness of lexical guidance and corrective feedback for L2 perceptual learning in the more natural setting of conversational interaction. Phonetic instruction, particularly perception-focused instruction that aims to direct listeners' attention to specific phonemes or phonetic cues, has primarily been studied for non-native sounds in unfamiliar languages (e.g., Chandrasekaran et al., 2016; Chen & Pederson, 2017; Guion & Pederson, 2007; Hisagi & Strange, 2011; Pederson & Guion-Anderson; Porretta & Tucker, 2014). Chapter 6 investigated phonetic instruction for younger and older adult L2 listeners who were already proficient in the non-native language, thereby extending prior research to a wider listener age range and to a more realistic learning context.

7.2.1 Implicit lexical guidance

Chapters 4 and 5 investigated whether implicit lexical guidance in a task-based dialogue led to perceptual learning of a vowel shift embedded in an unfamiliar L2 accent. Both chapters employed a comparable dialogue session in which, as part of a cooperative game, participants needed to interpret their interlocutor's /ɪ/ pronunciations as representing /ɛ/ so that the interlocutor's spoken words would match the onscreen lexical information. Each chapter tested participants' resultant perceptual learning in a different way: Chapter 4 used a lexical decision post-test, while Chapter 5 used a phonetic categorization pre-test and post-test.

The lexical decision post-test results of Chapter 4 showed that, as expected, participants in the lexical guidance condition were significantly faster to correctly respond "yes" to critical accented words (e.g., "best" pronounced */bɪst/), compared to participants in the control and corrective feedback conditions (to be discussed in Section 7.2.2). For the acceptance rate of the critical words, there was only a trend in the expected direction: the lexical guidance participants were numerically, but not significantly, more likely to accept these words than control or corrective feedback participants. However, the critical words' acceptance rate was significantly higher as a function of higher word identification accuracy, or "uptake," during the dialogue phase, which was by design at ceiling for the lexical guidance participants and at floor for the control participants. Thus, lexical guidance still impacted the critical words' acceptance rate indirectly, via its effect on participants' interpretation of their partner's pronunciations during the dialogue. The faster online processing for critical accented words aligns with the findings of Maye et al. (2008), who also found changes in listeners' online lexical processing consistent with their exposure to a cross-category vowel shift. Also, like in Maye et al.'s (2008) study, lexical guidance in Chapter 4 had no effect on listeners' online processing of items that would already sound like real words before exposure to the accent (e.g., *"geft" pronounced /gɪft/): such items were universally accepted as real words, with no significant reaction time differences between conditions. This shows that listeners did not learn that their interlocutor's /ɪ/ pronunciations *must* represent /ɛ/, but they learned that /ɪ/ *could* represent /ɛ/. Considering the very limited number of occurrences of accented words in the dialogue phase (16 critical words spread throughout an approximately 15-minute interaction), it is not surprising that listeners' perceptual adjustments would be rather conservative.

The phonetic categorization pre- and post-test in Chapter 5 shed light on how perception of the vowel shift, as measured by how much of the /ɛ/-/ɪ/ vowel continuum was perceived as /ɛ/, differed not only between participant groups but also within participant groups over time. If participants had adapted to the accent, learning that their interlocutor's /ɪ/ pronunciations represent /ɛ/, they should make more /ɛ/ categorization responses in the post-test. The effect of lexical guidance was clear when looking only at the post-test: as expected, participants in lexical guidance condition categorized significantly more of the continuum as /ɛ/ than participants in the control condition. However, the picture was more complicated when considering the pre-test and post-test responses together: in fact, participants in both the control and lexical guidance conditions made more /ɪ/ responses in the post-test than in the pre-test. Thus, the lexical guidance only attenuated the overall post-test shift toward /ɪ/. Post-hoc analyses ruled out that the unexpected post-test /ɪ/-shift might have been compensating for a pre-test /ɛ/-bias, or that responses within each test tended toward /ɪ/ in later trials. Therefore, it seems more likely that the additional exposure to the speaker's voice and her realization of other vowels during the intervening dialogue phase may have altered how listeners mapped her entire vowel space. This explanation reflects a possibility raised by McQueen and Mitterer (2005), who reported that lexically guided perceptual learning of vowels may have been affected by listeners' exposure to the same speaker's other vowel continua in other testing blocks, which led to presentation order effects across the blocks.

Taken together, Chapters 4 and 5 demonstrate the robustness of lexically guided perceptual learning, as the effects of lexical guidance on lexical and phonemic processing were detectable even within the relatively challenging context of L2 listening to a wholly unfamiliar accent in a task-based dialogue. This research thus extends the small set of studies that have examined lexically guided learning of vowels (Maye et al., 2008; Mitterer & McQueen, 2005) and lexically guided learning in L2 listening (Cooper & Bradlow, 2018; Drozdova et al., 2016).

Regarding the interactive context, Chapter 4 combined data from both the face-to-face ventriloquist paradigm and the computer-player setup (as described in Chapter 3) and found no significant learning differences between the two settings, even though the physically co-present interlocutor might have induced a higher cognitive processing load (see Sjerps et al., 2020). The robustness of the lexical guidance effects here thus support research showing that lexical guidance is effective under various types of cognitive load (Zhang & Samuel, 2014). Moreover, as participants in both of the interactive paradigms showed perceptual learning despite having to alternate speaking and listening throughout the task, these results contrast with previous studies showing disruptions to perceptual learning when listeners had to speak during training (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007). In the present studies, perhaps the speech production was not so disruptive because there was a relatively high degree of separation between production and perception, both in terms of the time between the participants' speaking and listening turns (at least several seconds) and in terms of the content of what speech participants heard (key words belonging to onscreen minimal pairs) and what speech they had to produce (descriptions of colored shapes needed to solve puzzles). Overall, the results fill an important gap in the literature by demonstrating that lexical guidance drives perceptual learning not only in passive listening but also in communicative interaction.

7.2.2 Interactional corrective feedback

The effect of interactional corrective feedback on uptake and online processing of a novel L2 accent was one of the main questions addressed in Chapter 4. During a task-based dialogue featuring puzzles and minimal word pairs, some participants received corrective feedback from their interlocutor on critical trials whenever they misperceived her accented pronunciation and mistakenly clicked on a wrong word as a result. Two types of corrective feedback were compared: generic corrective feedback, in which the interlocutor simply remarked that the answer was wrong (e.g., "Oh no, wrong one"), and contrastive corrective feedback, in which the interlocutor explicitly contrasted the misperceived word with the intended word (e.g., ""Oh oops, what you wanted was *pen*, not *pin*").

First, results showed that only participants receiving contrastive corrective feedback showed uptake for the accent, becoming more accurate at identifying critical words correctly over the course of the interaction. Participants receiving generic corrective feedback, in contrast, did not improve over time and maintained a near floor-level word identification accuracy, just like the control participants who never received feedback. The superiority of contrastive feedback over generic feedback is consistent with the results of Lee and Lyster's (2016b) computer-based L2 perceptual training study, in which auditory corrective feedback that contrasted the two members of a minimal pair was superior to visual corrective feedback consisting of the word "wrong." Furthermore, as the present study's contrastive feedback was more explicitly about perception, whereas the generic feedback left the source of the error implicit, these results support Rassaei's (2013) finding that explicit correction was more beneficial for L2 grammatical development than implicit recasts. Overall, the relatively low uptake observed for either type of corrective feedback, along with the substantial variability in individuals' responsiveness to it, suggests that the interpretability of the feedback in this experiment was not always straightforward. For instance, listeners might have mistakenly attributed the generic feedback to factors unrelated to their own perception, such as the puzzle being solved incorrectly or the interlocutor herself misspeaking. In line with Mackey et al.'s (2000) research, the accuracy of learners' perceptions about interactional feedback, particularly more implicit types of feedback, can therefore not be taken for granted.

Second, when it came to online processing of the novel L2 accent, results showed that there was no significant direct effect of corrective feedback condition: the performance of the contrastive and generic corrective feedback participant groups did not differ from that of the control group in the lexical decision post-test. However, similarly to lexical guidance (as discussed in Section 7.2.1), corrective feedback did appear to influence online processing indirectly: listeners who had exhibited more uptake for the accent during the interaction, as operationalized by their word identification accuracy, showed faster and more accurate online processing of critical accented words in the subsequent lexical decision test. In other words, lexical processing was faster and more accurate for those individuals who came to interpret the accent accurately during the interaction. This result supports a weak version of Schmidt's (2001) noticing hypothesis, in that listeners who accurately interpreted the interactive feedback, believing it to reflect a true mismatch between what they perceived and what their interlocutor actually said, were the ones whose online processing of the accent improved. While the present study thus demonstrated that interactional corrective feedback for speech perception can be effective, it also turned out that even relatively explicit, contrastive corrective feedback was less effective than implicit lexical guidance, which probably results at least in part from the fact that the corrective feedback could not be interpreted as reliably as the lexical information.

Finally, between the face-to-face ventriloquist paradigm and the computer-player interactive setting, there were no significant differences in either uptake or online processing due to corrective feedback. As the key issue was whether participants believed the feedback was really about their own perception, it appears that having a physically co-present interlocutor did not make the intention behind the feedback any clearer. Other factors that might have made the interactional corrective feedback even more effective and interpretable merit further investigation (see Section 7.4). One of the key contributions of Chapter 4 is that it extended the interactionist perspective on L2 acquisition (e.g., Ellis, 1999, 2003; Long, 1980) to the understudied domain of L2 speech perception, and at the same time, it extended prior training-based research on L2 speech perception (e.g., Bradlow et al., 1999; Iverson et al., 2005; Lee & Lyster, 2016b; Wang & Munrow, 2004) to the more natural learning context of conversational interaction.

7.2.3 Explicit phonetic instruction

The main aim of Chapter 6 was to investigate the effectiveness of explicit phonetic instruction for improving phonological awareness and perception of L2 sound contrasts in younger and older adults. Two English contrasts were examined: word-final /t/-/d/ (a familiar contrast in an unfamiliar position for native Dutch listeners, expected to be easier) and /æ/-/ɛ/ (a completely unfamiliar contrast to Dutch listeners, expected to be more difficult). First, supporting the theoretical link between conscious awareness and acquisition of specific L2 forms (Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994), the relationship between listeners' awareness and perception for the experimental contrasts at the outset of the experiment was established: at pre-test, there were moderate positive correlations between the Dutch listeners' awareness and perception for the English word-final /t/-/d/ contrast and between their awareness and perception for the English /æ/-/ɛ/ contrast.

Second, results showed that between the pre-test and post-test, phonological awareness generally increased more for the sound contrast that was featured in the video instruction, in line with previous perception studies showing that directing listeners' attention to a specific contrast increases perception of the attended contrast more than the non-attended contrast (Chen & Pederson, 2017; Pederson & Guion-Anderson, 2010). For phonological awareness-raising, the benefit of providing information about the phonetic cue of vowel duration differed as a function of age group. For both /t/-/d/ and /æ/-/ɛ/, young adults' awareness increased more after watching a duration-cue video than after a no-cue video (regardless of which sound contrast the video was about), suggesting that they could effectively transfer vowel length information about one contrast to reinforce their awareness of the other contrast. For the older adults, however, hearing about the vowel-cue information in the context of a consonant contrast was unhelpful for awareness-raising: at least for /t/-/d/, their awareness actually improved more after watching the /t/-/d/ no-cue video than after the /t/-/d/ duration-cue video. Furthermore, the older adults made fewer awareness gains overall. While the younger adults' awareness of both contrasts increased from pre-test to post-test in all conditions, suggesting that they may have also gained awareness simply by doing the intervening perception tests, the older adults' /æ/-/ɛ/ awareness only improved after watching an /æ/-/ɛ/ video. The fact that younger adults were more able than older adults to gain phonological awareness via generalization corroborates previous research showing that older adults exhibit fewer transfer effects in perceptual learning than younger adults (Bieber & Gordon-Salant, in press; Peelle & Wingfield, 2005).

Third, results showed that perceptual accuracy generally increased more for the sound contrast that was featured in the video instruction, though perceptual gains differed per age group and depending on whether duration cue information was provided. The younger adults' /t/-/d/ perception improved only after a /t-/d/ instruction, and likewise their /æ/-/ɛ/ perception improved only after an /æ/-/ɛ/ instruction, again consistent with previous research attesting to perceptual improvement specifically for contrasts that listeners were instructed to attend to (Chen & Pederson, 2017; Pederson & Guion-Anderson, 2010). The older adults' /t/-/d/ perception also improved only after a /t/-/d/ instruction, but perceptual improvement for /æ/-/ɛ/ was limited to those older adults who had received the /æ/-/ɛ/ duration-cue instruction. Thus, prior research about perceptual improvement from explicit instruction about a phonetic cue (Chandrasekaran et al., 2016; Hisagi & Strange, 2011; Porretta & Tucker, 2014) was supported but only partially, as the duration cue information did make a difference but only for the older adult listeners and for the more difficult of the two contrasts. Despite older listeners' potentially reduced perceptual flexibility (Scharenborg & Janse, 2013) and selective attention capacity (Sommers, 1997), they were apparently still able to focus on vowel duration as a cue to help them perceive the /æ/-/ɛ/ contrast, even though the vowel duration cue was detrimental to their awareness, at least for /t/-/d/. This suggests that the perceived relevance of instruction about a phonetic cue may be an important determiner for whether it benefits phonological awareness and perception.

Overall, Chapter 6 extended previous research about the acquisition of novel L2 sound contrasts by demonstrating how the relatively underutilized method of explicit phonetic instruction could improve phonological awareness and perception of both an easy and a more difficult L2 contrast. By implementing a short, one-time video instruction rather than a more traditional, time-intensive high-variability phonetic training method (Sakai & Moorman, 2018), and by focusing on already-proficient L2 listeners rather than those to whom the language is completely unfamiliar (as in Chandrasekaran et al., 2016; Chen & Pederson, 2017; Hisagi & Strange, 2011; Pederson & Guion-Anderson, 2010; Porretta & Tucker, 2014), this study aimed to represent a more natural language learning situation. Moreover, it built on just a handful of studies that have investigated L2 perceptual category learning in older adult listeners (Ingvalson et al., 2017; Kubo & Asahane-Yamada, 2006; Maddox et al., 2013; Tajima et al., 2002), highlighting differences between how younger and older adults benefit from certain aspects of the instruction.

7.3 Recommendations for L2 pedagogy

The present research about learning mechanisms for improving L2 speech perception leads to several pedagogical recommendations. The focus of this research was on improving L2 listeners' word recognition ability, which can involve learning to discriminate non-native sound contrasts or adapting to an unfamiliar L2 accent. First, Chapters 4 and 5 showed that implicit lexical guidance can improve L2 listeners' online lexical processing of an unfamiliar accent. Therefore, exposing L2 listeners to speech that matches their level of vocabulary knowledge, and providing visual lexical support simultaneously, may enable them to use the available lexical information to constrain their interpretation of the phonetic input, adapting their phonemic perception to the accent in question. While it is hardly feasible in real life to obtain visual lexical guidance for speech during real-time conversation, as in the present experiments, lexical guidance can be provided for plenty of naturalistic speech material for L2 listeners, for instance, in the form of foreign-language subtitles (Mitterer & McQueen, 2009).

Even though Chapter 4 showed implicit lexical guidance to be more effective than corrective feedback in conversational interaction, explicit information can also promote perceptual learning. Specifically, during the dialogue with an accented speaker, contrastive corrective feedback led to improved word recognition ability for some learners over the course of the interaction, while generic corrective feedback did not. Thus, in line with previous research about the benefits of explicit over implicit corrective feedback (Rassaei, 2013) and explicit over implicit instruction (Goo et al., 2015) for L2 learning (primarily of grammar), a good practice for L2 sound learning in the language classroom would be to provide L2 listeners with explicit feedback about their perceptual errors, making sure that the intention behind the feedback (correcting a misperception on the part of the listener) is unambiguous. Corrective feedback that contrasts the intended word with the misperceived word (e.g., "Pen, not pin") has the benefit of not only being easy to interpret but also providing additional exposure to the target form and making the contrast between the target and non-target form more available to short-term memory. While this recommendation is based on an experiment that used a relatively form-focused dialogue, making sure that interactional feedback is very clearly about the listener's own mistaken perception is probably even more crucial for learning in more free-ranging, meaning-focused communicative situations.

Finally, Chapter 6 showed that explicit phonetic instruction can both raise awareness of and improve perception of non-native sound contrasts in young adult and older adult L2 listeners. As Kissling (2012, 2014) discusses, this type of relatively short, explicit instruction is more applicable to and typical of L2 teaching settings than the intensive, implicit training programs traditionally used in perception research. Based on the findings in Chapter 6, a promising method for improving L2 phonological awareness and perception is a simple video-based instruction by a native speaker that draws learners' attention to minimal pairs based on difficult L2 contrasts. When the sound contrast is particularly difficult to discriminate (as when the two sounds assimilate to a single phonemic category in the L1, or in the case of older listeners with lower L2 proficiency), it can be beneficial for perceptual learning to provide information about a specific phonetic cue on which listeners can focus their attention. However, the relevance of information about specific phonetic cues to what is being taught or tested should also be taken into account, as certain less intuitive cues (e.g., the preceding vowel's duration as a cue to a final consonant contrast) might not benefit all learners for all language tasks.

7.4 Directions for future research

One interesting project for future research based on this dissertation would be to expand upon the present work on L2 perceptual learning in interaction. For starters, it would be useful to reproduce the basic design of the experiments in Chapters 4 and 5 but with a different accent for the interlocutor, as the combination of the chosen vowel shift (/ɛ/ to /ɪ/) and unfamiliar dialect (Middlesbrough English) in this dissertation may have damped the observed learning effects. The results of Chapter 4 showed that listeners' uptake for the unfamiliar accent in response to corrective feedback was lower than expected: there was almost no uptake in response to generic feedback, and even with the more effective contrastive corrective feedback, listeners' overall accuracy on the cooperative game's critical trials was only 16.2%, with only around a quarter to a third of listeners responding correctly by the end of the game. Furthermore, Chapter 5 showed that listeners in both the control and lexical guidance conditions actually shifted their phonemic boundaries opposite to the expected direction from pre-test to post-test. As in Mitterer and McQueen's (2005) study on lexically guided perceptual learning of vowels, it seems possible that exposure to the speaker's other (non-shifted) vowels may have affected the perceived applicability of what was learned about the critical vowel contrast. Perhaps the particular combination of the vowel shift and the carrier accent used in Chapters 4 and 5 was not very plausible, leading participants to disbelieve that the corrective feedback could be about their perception. Thus, future research based on a similar design might do well to give the interlocutor a naturally occurring dialect (whether a "standard" or unfamiliar one). This would entail that some participants might bring prior knowledge of the accent to the experiment, but this potential concern could be mitigated by testing lower-proficiency L2 learners than the ones in the present research. Moreover, particularly for listeners with much lower L2 proficiency, it would be interesting to investigate whether interactional corrective feedback and lexical guidance could promote perceptual learning for a non-native phonemic contrast that tends to be assimilated into a single L1 phonemic category, in line with Lee and Lyster's (2016a, 2016b) classroom and computer-based training studies.

In addition, there are many opportunities for research applying the ventriloquist paradigm to new research questions about phonological learning and alignment in interaction. For instance, there are still questions about how L2 sound learning from interactional feedback might be affected by factors that the present experiments did not manipulate. One challenge in designing the dialogue task in Chapter 4 was determining the appropriate amount of interactional feedback, or critical trials, to include: too little feedback might be insufficient for learning, but too much feedback might come across as unnatural or confrontational. Using multiple dialogue sessions, rather than a single dialogue, would make it easier to systematically vary the amount and timing of feedback provided, as well as enabling tests of the long-term retention of any resultant learning. The target of interactional feedback can also be manipulated. While the present research focused on providing L2 listeners with feedback on their *perception*, the ventriloquist paradigm has also recently been used to study feedback on their *pronunciation*: specifically, whether L2 speakers' pronunciation and perception of novel L2 sound contrasts improves as a result of implicit negative feedback about their pronunciation provided by an L1 interlocutor (Troncoso-Ruiz et al., 2019; Ye, 2020). If indeed robust L2 sound learning from interactional feedback can be demonstrated, then follow-up studies based on communication accommodation theory (Giles, 1973; Giles & Ogay, 2007; Shepard et al., 2001) could manipulate psychosocial and sociolinguistic factors that might impact the amount of learning from dialogue, such as the status of the interlocutor as an authority on the language (e.g., being a teacher vs. student or being a native vs. non-native speaker; see Llurda, 2005) or other aspects of the interlocutor's social identity, such as gender or dialect, that have been shown to play a role in phonetic alignment (e.g., Babel et al., 2014; Pardo, 2006). More broadly, the ventriloquist paradigm could be used not only for studying phonetic learning in a second language but also for studying how any of the above factors affect phonetic alignment between native speakers in conversational interaction.

7.5 Conclusions

The goal of this dissertation was to provide a methodological basis for studying second language speech perception in natural settings and to investigate how several different learning mechanisms, ranging from implicit to explicit, might improve L2 speech perception.

Two methodological innovations were developed, described, and validated. The dictation task (Chapter 2) was shown to be a valuable tool for studying the perception of conversational speech when transcriptions are scored with a range of measures that provide more detailed information about several aspects of the input listeners can recover. In addition, moving beyond the study of passive listening to the study of active conversational interaction, the newly developed ventriloquist paradigm (Chapter 3) was presented as an invaluable method for studying speech processing in a natural dialogue context, reconciling ecological validity with experimental control over phonetic input.

In the remaining chapters, three learning mechanisms for L2 speech perception were investigated: lexical guidance, interactional corrective feedback, and explicit phonetic instruction. First, implicit lexical guidance was shown to promote perceptual learning of a novel accent in dialogue, demonstrating for the first time that this learning mechanism works even in the cognitively demanding setting of L2 listening in a communicative task (Chapters 4 and 5). Second, interactional corrective feedback was demonstrated to promote uptake for L2 speech perception (Chapter 4), in particular when the feedback emphasized the contrast between what the listener thought they heard and what their interlocutor actually said. Third, explicit phonetic instruction featuring minimal word pairs was shown to improve phonological awareness and perception of two L2 contrasts in both younger adult and older adult listeners (Chapter 5), providing evidence for the benefit of directing L2 listeners' conscious attention to specific sounds and to phonetic cues.

Taken together, the research in this dissertation shows that combining insights from research in the fields of phonetics, psycholinguistics, and second language acquisition can improve our understanding of second language speech perception learning in natural contexts.

Nederlandse samenvatting[4]

Leren luisteren in een tweede taal is een cruciale component van tweede taalverwerving en speelt een belangrijke rol in succesvolle communicatie. Spraakperceptie in een moedertaal (L1) is meestal een moeiteloos proces, maar spraakperceptie in een tweede taal (L2) is vaak lastiger omdat het bijvoorbeeld moeilijker is om woorden te herkennen (Cutler, 2012). Eén van de uitdagingen is het verschil kunnen horen tussen twee spraakklanken die gedifferentieerd dienen te worden in de L2 maar niet in de L1 (Best, 1994; Best & Tyler, 2007). Het komt bijvoorbeeld vaak voor dat Nederlandse luisteraars de Engelse woorden "pan" (/pæn/) en "pen" (/pɛn/) door elkaar halen, omdat de fonemen /æ/ en /ɛ/ perceptueel geassimileerd worden, en worden waargenomen als één fonetische categorie in hun L1. Een tweede uitdaging voor L2 luisteraars is om hun spraakperceptie aan te passen aan regionale of buitenlandse accenten. Een Nederlandse reiziger in Australië zou bijvoorbeeld de woorden "pen" (/pɛn/) en "pin" (/pɪn/) door elkaar kunnen halen omdat hij niet bekend is met de /ɛ/-naar-/ɪ/ klankverschuiving in sommige Australisch Engelse dialecten. Een belangrijke vraag in tweede taalverwerving is hoe L2 luisteraars deze uitdagingen van woordherkenning kunnen overkomen om hun perceptie van L2 spraak te verbeteren.

L2 spraakverwerking en perceptueel leren zijn traditioneel onderzocht met niet-interactieve computerprogramma's die intensieve blootstelling bieden aan sterk gecontroleerde stimuli, zonder enige context (e.g., Bradlow et al., 1999; Iverson & Evans, 2009; Lively et al., 1994; Logan et al., 1991; Sakai & Moorman, 2018). Deze experimentele paradigma's hebben waardevolle inzichten opgeleverd over hoe verschillende aspecten van de fonetische blootstelling bijdragen aan perceptueel leren. Echter, ze omvatten niet de natuurlijkere leersituaties die L2 luisteraars tegenkomen in het dagelijkse leven, zoals het leren door middel van interactie. Het combineren van controle over de fonetische blootstelling met ecologische validiteit is een belangrijk methodologisch doel voor onderzoek naar het leren van L2 spraak: dit zou het

[4] Thanks to Tim Zee for proofreading this chapter thoroughly and making numerous language-related corrections and improvements.

mogelijk maken om te testen in hoeverre bepaalde leermechanismes voor L2 spraak, die voorgesteld zijn op basis van onderzoek met voornamelijk onnatuurlijke luistercontexten, ook van toepassing zijn op alledaagse communicatieve situaties.

Dit proefschrift ging over het leren van L2 spraakperceptie in natuurlijke situaties, zoals in gesprekken of door middel van eenvoudige fonetische instructie. Het was in het bijzonder gericht op de vraag hoe L2 luisteraars hun woordherkenningsvermogen kunnen verbeteren door zich aan te passen aan een onbekend accent en door het leren onderscheiden van een klankcontrast in de L2 dat niet bestaat in de L1. Het proefschrift had twee doelen: ten eerste, het presenteren en evalueren van twee ecologisch valide methoden voor het onderzoeken van spraakverwerking, en ten tweede, het nader onderzoeken van drie leermechanismes voor L2 spraakperceptie: impliciete lexicale begeleiding, interactieve corrigerende feedback en expliciete fonetische instructie.

Methodologische bijdragen aan onderzoek naar meer natuurlijke taalverwerking

Onderzoekers in de psycholinguïstiek en gerelateerde domeinen hebben de afgelopen jaren opgeroepen tot het gebruik van meer ecologisch valide methoden in taalverwerkingsonderzoek, bijvoorbeeld door het gebruik van natuurlijkere spraakstimuli, zoals continue spraak uit informele gesprekken, en door een focus op natuurlijkere situaties voor taalgebruik, zoals interactieve dialogen (Tanenhaus & Brown-Schmidt, 2008; Tucker & Ernestus 2016; Willems 2017). Het dictee, een taak waarin leerlingen proberen uitgesproken zinnen foutloos op te schrijven, bevat wel continue spraak en wordt vaak gebruikt in de context van L2 taalonderwijs (Buck, 2001; Oller & Streiff, 1975; Savignon, 1982; Stansfield, 1985). Echter, de traditionele beoordelingsmethoden voor een dicteetaak, die gebaseerd zijn op alleen de hoeveelheid juiste woorden in een antwoord, geven een incompleet beeld van de fonetische en semantische aspecten van de taalinput die luisteraars kunnen herkennen. Bovendien, hoewel interactieve gesprekken een belangrijke leercontext zijn voor algemene L2 verwerving (Ellis, 1999, 2003; Long, 1980), wordt het leren van L2 fonologie bijna nooit onderzocht binnen interactieve dialogen omdat het daarbij moeilijk is om de fonetische blootstelling te controleren. Hoofdstuk 2 van dit proefschrift toonde aan hoe het dictee een waardevoller onderzoeksinstrument voor de taalwetenschap kon worden door het toepassen van objectievere en informatievere beoordelingsmethoden. Verder introduceerde Hoofdstuk 3 een geheel nieuw experimenteel paradigma voor het bestuderen van spraakverwerking in dialogen. Het "buiksprekersparadigma" combineert een overtuigende illusie van een live gesprek tussen een proefpersoon en een onderzoeker met controle over de fonetische input van de proefpersonen dankzij het gebruik van vooraf opgenomen spraak.

Onderzoek naar de perceptie van informele continue spraak met dictee

Hoofdstuk 2 beschreef en evalueerde vier verschillende maten voor het beoordelen van de transcripties van een dictee, in dit geval een dictee waarin L1 en L2 luisteraars enkele korte zinnen uit informele gesprekken moesten transcriberen. De volgende vier beoordelingsmaten werden beschreven en geëvalueerd: lexicale foutenpercentage, orthografische bewerkingsafstand, fonologische bewerkingsafstand en semantische foutenpercentage. Ten eerste werd de discriminante validiteit van de maten ondersteund door resultaten die lieten zien dat de L1 luisteraars significant beter presteerden dan de L2 luisteraars in alle vier de maten; met deze maten konden de twee groepen luisteraars dus goed onderscheiden worden. Dit verschil was verwacht omdat luisteren in een L2 over het algemeen moeilijker is dan luisteren in een L1 (Cutler, 2012) en omdat eerder onderzoek heeft aangetoond dat gereduceerde uitspraakvarianten, die aanwezig waren in de zinnen in dit experiment, bijzonder moeilijk zijn voor L2 luisteraars (bv. Brand & Ernestus, 2018; Ernestus et al., 2017). Ten tweede werd aangetoond dat de vier maten van elkaar verschilden in het kwantificeren van luisterprestatie: de luisterprestatie was het slechtst volgens het lexicale foutenpercentage, beter volgens het semantische foutenpercentage en de orthografische bewerkingsafstand, en het beste volgens de fonologische bewerkingsafstand. Dit suggereert dat het lexicale foutenpercentage, de maat die traditioneel gebruikt wordt in de context van vreemde talenonderwijs (Buck, 2001), de luistervaardigheid mogelijk onderschat als het gaat om het vermogen van luisteraars om de betekenis en de klanken van spraak te achterhalen. Ten derde werd de criteriumvaliditeit van de maten onderbouwd door de bevinding dat de vier maten significant gecorreleerd waren met de door de luisteraars zelf beoordeelde L2 vaardigheid, het aantal uren dat ze naar Engels luisterden (maar niet spraken) per week, en hun L2 woordkennis zoals objectief gemeten door LexTALE (Lemhöfer & Broersma, 2012). Tot slot werd aangetoond dat de vier maten sterk gecorreleerd waren met elkaar, met de laagste correlatie tussen de semantische en fonologische maten zoals verwacht. Het hoofdstuk concludeerde met een discussie over welke combinaties van maten het beste gebruikt kunnen worden voor onderzoek naar spraakperceptie op basis van praktische en theoretische overwegingen.

Onderzoek naar spraakverwerking in interactie met het buiksprekersparadigma

Hoofdstuk 3 presenteerde en valideerde een recent ontwikkelde experimentele methode, het buiksprekersparadigma, dat de mogelijkheid biedt om spraakverwerking in interactie te onderzoeken met volledige controle over de fonetische blootstelling aan de proefpersonen. In deze methode voeren proefpersonen een één-op-één, face-to-face gesprek met de onderzoeker, die communiceert (buiten het medeweten van de proefpersoon) door het afspelen van eerder opgenomen spraakopnames. De spraakopnames bevatten zowel taak-gerelateerde uitdrukkingen als flexibele uitdrukkingen van verschillende categorieën, zoals algemene bevestigende en ontkennende antwoorden, beweringen zoals “ik weet het niet,” verzoeken om opheldering of herhaling en korte reacties om begrip en aandacht te signaleren. De spraakopnames worden afgespeeld in de kooptelefoon van de proefpersoon, wat ogenschijnlijk komt doordat de gesprekspartners communiceren door te spreken in de microfoons op tafel. Iedere keer dat de onderzoeker “spreekt” duikt ze even achter haar beeldscherm en gebruikt ze een verborgen toetsenbord om de gewilde spraakopname af te spelen.

Ten eerste werd de validiteit van dit paradigma vastgelegd door analyses van ongeveer honderd experimentele sessies, die gebruik maakten van twee verschillende gesprekspartners en twee verschillende opgenomen sprekers. Ongeveer tachtig procent van de proefpersonen rapporteerde geen enkel vermoeden te hebben dat de spraak van hun gesprekspartners vooraf opgenomen was; zij waren er dus waarschijnlijk van overtuigd dat ze een normaal gesprek aan het voeren waren. Ten tweede bewezen de analyses het belang van de face-to-face setting om het gesprek geloofwaardig te maken: in een alternatieve setting waar de proefpersoon en onderzoeker in aparte testcabines zaten, geloofden maar één derde van de proefpersonen dat hun gesprekspartner een echte persoon was, terwijl de overige proefpersonen geloofden dat zij en computer of robot was. De face-to-face setting blijkt dus cruciaal te zijn om de illusie dat het gesprek authentiek was te onderhouden. Ten derde liet dit hoofdstuk zien dat het buiksprekersparadigma leidde tot een levendiger gesprek (zoals gemeten in het spraakgedrag van de proefpersonen) dan een vergelijkbaar paradigma waarin proefpersonen verteld werden dat hun gesprekspartner alleen een computer was.

Het buiksprekersparadigma vereist zorgvuldige voorbereiding om de interactie natuurlijk en geloofwaardig te maken, en deze geloofwaardigheid is fundamenteel voor de ecologische validiteit. Verschillende manieren om de natuurlijkheid en geloofwaardigheid van het paradigma nog verder te verbeteren, kort beschreven in Hoofdstuk 7, werden geïmplementeerd in een nieuw experiment dat ik begeleidde, waarin rond de negentig procent van de proefpersonen ervan overtuigd was dat de interactie authentiek was (Ye, 2000). Hoewel het buiksprekersparadigma wellicht nooit honderd procent van de proefpersonen zal kunnen overtuigen, zijn de succesvolle aanpassingen van Ye (2020) een veelbelovend teken dat de natuurlijkheid en geloofwaardigheid van het paradigma continu verbeterd kunnen worden.

Leermechanismes voor L2 spraakperceptie

Dit proefschrift onderzocht drie verschillende leermechanismes voor L2 spraakperceptie: impliciete lexicale begeleiding, interactieve correctieve feedback en expliciete fonetische instructie. Lexicale begeleiding wordt meestal onderzocht in niet-interactieve luistersituaties (zie de reviews van Samuel & Kraljic, 2009 en Baese-Berk, 2018), en evenzo wordt correctieve feedback voor spraakperceptie meestal onderzocht met intensieve computertrainingsprogramma's (bv. Bradlow et al., 1999; Iverson et al., 2005; Lee & Lyster, 2016b; Wang & Munrow, 2004). Hoofdstukken 4 en 5 onderzochten de effectiviteit van lexicale begeleiding en correctieve feedback voor het leren van L2 spraakperceptie in de meer natuurlijke context van een interactief gesprek. Fonetische instructie, met name instructie voor perceptie die de aandacht van luisteraars vestigt op specifieke fonemen of fonetische details zoals klinkerlengte, wordt meestal onderzocht met klanken en talen die helemaal onbekend zijn voor de luisteraars (bv. Chandrasekaran et al., 2016; Chen & Pederson, 2017; Guion & Pederson, 2007; Hisagi & Strange, 2011; Pederson & Guion-Anderson; Porretta & Tucker, 2014). Hoofdstuk 6 onderzocht het effect van fonetische instructie voor zowel jongere als oudere volwassen L2 luisteraars die al vaardig waren in de vreemde taal; daarmee werd eerder onderzoek in dit domein uitgebreid naar meer leeftijdsgroepen en tegelijkertijd naar een realistischere leercontext.

Impliciete lexicale begeleiding

Hoofdstukken 4 en 5 onderzochten of impliciete lexicale begeleiding in een taakgerichte dialoog perceptueel leren van een klinkerverschuiving in een onbekend L2 accent bevordert. Beide hoofdstukken gebruikten een vergelijkbare dialoogsessie met een coöperatief computerspel. De proefpersonen moesten de /ɪ/-uitspraken van hun gesprekspartner leren interpreteren als /ɛ/ zodat de door de gesprekspartner uitgesproken woorden overeen zouden komen met de lexicale informatie op het beeldscherm. De hoofdstukken testten het resulterende perceptueel leerproces op twee manieren. Hoofdstuk 4 gebruikte een lexicale decisie post-test, waarin luisteraars moesten aangeven of bepaalde uitspraken van hun partner wel of niet bestaande woorden waren. Uit deze test moest blijken of de lexicale begeleiding een effect had op het proces van woordherkenning. Hoofdstuk 5 gebruikte een fonetische categorisatie pre-test en post-test, waarin luisteraars moesten aangeven of bepaalde ambigue uitspraken, waarvan de klinkers gemanipuleerd waren om ergens tussen een pure /ɪ/ en /ɛ/ te vallen, klonken als /ɪ/-woorden of als /ɛ/-woorden. Op deze manier kon de perceptuele grens tussen /ɪ/ en /ɛ/ vastgelegd worden om te zien of de grens veranderde van pre-test tot post-test.

De lexicale decisie post-test resultaten van Hoofdstuk 4 lieten zoals verwacht zien dat proefpersonen in de lexicale begeleiding conditie significant sneller waren om correct "ja, dat is een woord" te antwoorden op kritische woorden met de klinkerverschuiving (bijvoorbeeld "best" uitgesproken als het niet-woord "bist" */bɪst/), vergeleken met proefpersonen in de controle en correctieve feedback condities. Naast het significante effect in de reactietijden was er een niet-significante trend in de verwachtte richting voor de acceptatie van deze kritische woorden: proefpersonen in de lexicale begeleiding conditie antwoordden (numeriek maar niet statistisch significant) vaker "ja" op deze woorden dan proefpersonen in de andere condities. Niet alleen de experimentele conditie zelf maar ook het leergedrag tijdens het interactieve computerspel had een significant effect in de lexicale decisie taak: hoe nauwkeuriger de proefpersonen de klinkerverschuiving geleerd hadden tijdens de interactie, hoe vaker ze daarna "ja" antwoordden op kritische lexicale decisie woorden. Het experiment was zo ontworpen dat de nauwkeurigheid op kritische woorden tijdens de dialoog vrijwel nul moest zijn in de controle conditie en vrijwel honderd procent moest zijn in de lexicale begeleiding conditie. Deze laatste bevinding laat dus zien dat lexicale begeleiding wel een indirect effect had op hoe vaak kritische woorden geaccepteerd werden in de lexicale decisie taak, middels het effect op de interpretatie van uitspraken tijdens de interactie.

De fonetische categorisatie pre- en post-test in Hoofdstuk 5 brachten aan het licht dat perceptie van de klinkerverschuiving, geoperationaliseerd als hoeveel van het /ɛ/-/ɪ/ klinkercontinuüm werd waargenomen als /ɛ/, niet alleen tussen proefpersonen verschilde maar ook binnen proefpersonen in de loop van de tijd veranderde (voor en na het dialoog). Als proefpersonen hun perceptie hadden aangepast aan het accent, en dus geleerd hadden dat de /ɪ/-uitspraken van hun gesprekspartner eigenlijk als /ɛ/ geïnterpreteerd moesten worden, dan zouden ze meer /ɛ/ categorisatie reacties moeten geven in de post-test. Het effect van lexicale begeleiding was duidelijk als alleen de post-test werd beschouwd: zoals verwacht categoriseerden proefpersonen in de lexicale begeleiding conditie meer van het klinkercontinuüm als /ɛ/ dan proefpersonen in de controle conditie. Echter, het beeld werd ingewikkelder als de resultaten van de pre-test en post-test samen beschouwd werden, want eigenlijk gaven proefpersonen in beide condities meer /ɪ/-antwoorden in de post-test dan in de pre-test. De lexicale begeleiding verzwakte dus blijkbaar alleen de algemene verschuiving naar /ɪ/ in de post-test. Een mogelijke verklaring voor deze onverwachte /ɪ/-verschuiving is dat de blootstelling aan de stem van de spreker en haar uitspraak van andere klinkers tijdens het dialoog perceptuele veranderingen heeft veroorzaakt in hoe luisteraars het gehele klinkersysteem van de spreker indeelden.

Samen genomen ondersteunen Hoofdstukken 4 en 5 de robuustheid van lexicaal gedreven perceptueel leren, want de effecten van lexicale begeleiding op lexicale en fonemische verwerking waren zelfs detecteerbaar binnen de relatief uitdagende luistercontext van een onbekend L2 accent in een taakgerichte dialoog. Dit onderzoek bouwt voort op het kleine aantal studies over het lexicaal gedreven leren van klinkers (Maye et al., 2008; Mitterer & McQueen, 2005) en het lexicaal gedreven leren in L2 perceptie (Cooper & Bradlow, 2018; Drozdova et al., 2016). Met betrekking tot de interactieve leercontext combineerde Hoofdstuk 4 data van zowel het face-to-face buiksprekersparadigma als de niet-face-to-face computer-speler versie van het experiment (zoals beschreven in Hoofdstuk 3). Er werden geen significante verschillen tussen de twee versies gevonden wat betreft perceptueel leren, ook al zou de fysieke aanwezigheid van de gesprekspartner misschien geleid hebben tot een hogere cognitieve belasting (zie Sjerps et al., 2020). De robuustheid van de effecten van lexicale begeleiding ondersteunen dus voorgaand onderzoek dat liet zien dat lexicale begeleiding effectief is onder verschillende vormen van cognitieve belasting (Zhang & Samuel, 2014). Al met al tonen deze resultaten aan dat lexicale begeleiding het perceptueel leren van spraakklanken niet alleen bevordert tijdens passief luisteren maar ook in communicatieve interactie.

Interactieve correctieve feedback

In Hoofdstuk 4 was één van de hoofdonderzoeksvragen wat het effect van interactieve correctieve feedback is op *uptake* (begrip) en online verwerking van een onbekend L2 accent. Tijdens een taakgerichte dialoog, waarin puzzels en minimale woordenparen centraal stonden, kregen sommige proefpersonen correctieve feedback van hun gesprekspartner op kritische momenten, namelijk iedere keer dat ze haar accent verkeerd begrepen en daardoor op een verkeerd woord klikten. Twee soorten correctieve feedback werden vergeleken: generieke correctieve feedback, waarbij de gesprekspartner de simpele opmerking maakte dat het antwoord fout was (bijvoorbeeld "Oh no, wrong one"), en contrastieve correctieve feedback, waarbij de gesprekspartner een expliciet contrast maakte tussen het woord dat de proefpersoon dacht te horen en het woord dat bedoeld was (bijvoorbeeld "Oh oops, what you wanted was *pen*, not *pin*").

Ten eerste lieten proefpersonen alleen *uptake* zien in de contrastieve feedback conditie: ze werden beter in het identificeren van kritische woorden gedurende de interactie. Proefpersonen in de generieke feedback conditie, daarentegen, begrepen het accent niet beter in de loop van de interactie en konden de kritische woorden met de klinkerverschuiving vrijwel nooit correct identificeren, net zoals de controle proefpersonen die helemaal geen feedback ontvingen. Ten tweede, wat betreft de online verwerking van het accent was er geen direct effect van de correctieve feedback conditie: geen van de twee correctieve feedback groepen had een significant hogere score dan de controle groep in de lexicale decisie post-test. Echter, net als met lexicale begeleiding was er wel een indirect effect van correctieve feedback op online verwerking: luisteraars die meer *uptake* hadden getoond tijdens de interactie, geoperationaliseerd als een nauwkeurigere identificatie van de kritische woorden, gaven sneller en nauwkeuriger antwoorden op de kritische woorden in de lexicale decisie test. Met andere woorden, de automatische lexicale verwerking werd sneller en nauwkeuriger bij de luisteraars die het accent beter leerden begrijpen tijdens de interactie. Hoewel de huidige resultaten hebben aangetoond dat interactieve correctieve feedback online spraakverwerking kan verbeteren, bleek ook dat zelfs relatief expliciete, contrastieve feedback minder effectief was dan impliciete lexicale begeleiding. Dit komt waarschijnlijk deels door het feit dat de lexicale informatie beter geïnterpreteerd kon worden dan de correctieve feedback. Ten derde, tussen de buikspreker- en computer-speler versies van het experiment bleken er geen significante verschillen te zijn in hoeverre de correctieve feedback het leren van het accent, gemeten in *uptake* of online verwerking, bevorderde.

Expliciete fonetische instructie

Hoofdstuk 6 onderzocht de effectiviteit van expliciete fonetische instructie voor de verbetering van fonologisch bewustzijn en perceptie van L2 klankcontrasten in jongere en oudere volwassenen. Twee Engelse contrasten werden onderzocht: /t/-/d/ op woordfinale positie (een bekend contrast in een onbekende positie voor Nederlandse luisteraars, wat makkelijker zou kunnen zijn om te leren) en /æ/-/ɛ/ (een geheel onbekend contrast voor Nederlandse luisteraars, wat moeilijker zou kunnen zijn om te leren). Tussen pre-tests en post-tests die bewustzijn en perceptie maten keken Nederlandse proefpersonen naar een kort instructiefilmpje dat ging over /t/-/d/ of /æ/-/ɛ/ en dat wel of niet beschreef hoe verschillen in klinkerduur (de tijdsduur van /æ/-/ɛ/, of de tijdsduur van de klinker vóór /t/-/d/) gebruikt zouden kunnen worden om de twee klanken in het klankpaar te onderscheiden. Ten eerste werd de relatie tussen bewustzijn en perceptie van de kritische contrasten aan het begin van het experiment vastgesteld: in de pre-tests waren er matige positieve correlaties tussen bewustzijn en perceptie van de luisteraars bij zowel het woord-finale /t/-/d/ contrast als het /æ/-/ɛ/ contrast. Deze bevinding komt overeen met de theoretische link tussen bewustzijn en verwerving van specifieke L2 vormen (Schmidt, 1990; Svalberg, 2007; Tomlin & Villa, 1994).

Ten tweede is gebleken dat de toename in fonologisch bewustzijn tussen de pre-test en post-test groter was voor het klankcontrast dat uitgelegd werd in het instructieve filmpje dan voor het andere, niet-uitgelegde contrast. Dit resultaat voor fonologisch bewustzijn is in lijn met eerdere studies die aantoonden dat het richten van de aandacht op een specifiek contrast de perceptie van dat contrast meer verbeterd dan de perceptie van niet-belichte contrasten (Chen & Pederson, 2017; Pederson & Guion-Anderson, 2010). De toegevoegde waarde van de informatie over klinkerduur voor de verbetering van fonologisch bewustzijn verschilde tussen de twee leeftijdsgroepen. Voor zowel /t/-/d/ als /æ/-/ɛ/ werd het bewustzijn van jongere volwassen groter na het kijken van een filmpje met klinkerduurinformatie dan na het kijken van een filmpje zonder klinkerduurinformatie, ongeacht of het een /t/-/d/ of /æ/-/ɛ/

instructie was. Dit suggereert dat de jonge volwassenen de klinkerduurinformatie konden overbrengen van het ene contrast naar het andere om hun bewustzijn van het andere contrast te versterken. Voor oudere volwassenen was de klinkerduurinformatie in de context van het medeklinkercontrast echter niet nuttig voor het verbeteren van fonologisch bewustzijn. Bovendien verbeterde het bewustzijn van oudere volwassenen minder vaak dan het bewustzijn van jongere volwassenen: de jongere volwassenen werden meer bewust in de post-test voor beide klankcontrasten in alle condities, terwijl oudere volwassenen alleen hun /æ/-/ɛ/ bewustzijn verbeterden na een /æ/-/ɛ/ (dus niet na een /t/-/d/) instructie.

Ten derde bleek dat over het algemeen perceptuele nauwkeurigheid meer verbeterde voor het contrast dat werd uitgelegd in de instructie dan voor het niet-uitgelegde contrast. De perceptuele verbeteringen waren afhankelijk van de leeftijdsgroep en of de instructie wel of niet de klinkerduurinformatie bevatte. Voor jongere volwassenen verbeterde /t/-/d/ perceptie alleen na een /t/-/d/ instructie, en /æ/-/ɛ/ perceptie alleen na een /æ/-/ɛ/ instructie, in overeenkomst met voorgaand onderzoek waarin perceptuele verbetering alleen werd bewezen voor contrasten waar de aandacht van luisteraars op gericht werd (Chen & Pederson, 2017; Pederson & Guion-Anderson, 2010). Voor oudere volwassenen verbeterde /t/-/d/ perceptie ook alleen na een /t/-/d/ instructie, maar /æ/-/ɛ/ perceptie verbeterde alleen na de /æ/-/ɛ/ instructie met de klinkerduurinformatie. Deze informatie bleek wat betreft perceptuele accuratesse dus alleen belangrijk voor de oudere luisteraars en voor de moeilijkste van de twee contrasten. De effectiviteit van expliciete instructie over een fonetische kenmerk voor het verbeteren van L2 spraakperceptie, wat eerder onderzoek wel bewezen heeft (Chandrasekaran et al., 2016; Hisagi & Strange, 2011; Porretta & Tucker, 2014), werd in het huidige onderzoek dus alleen deels ondersteund.

Al met al bouwde Hoofdstuk 6 voort op voorgaand onderzoek over het verwerven van nieuwe L2 klankcontrasten door te bewijzen dat expliciete fonetische instructie, een relatief onderbenutte methode, verbeteringen kan veroorzaken in zowel fonologisch bewustzijn als perceptie van zowel een makkelijker als een moeilijker L2 klankcontrast. Deze studie poogde een natuurlijkere leersituatie te creëren door het gebruik van een korte, eenmalige video instructie in plaats van een lang en intensief traject van hoge-variabiliteit fonetische training (zie Sakai & Moorman, 2018) en door een focus op L2 luisteraars die al vaardig waren in de tweede taal in plaats van luisteraars voor wie de taal helemaal onbekend was (zoals in Chandrasekaran et al., 2016; Chen

& Pederson, 2017; Hisagi & Strange, 2011; Pederson & Guion-Anderson, 2010; Porretta & Tucker, 2014). Het draagt ook bij aan het kleine aantal voorgaande studies over het leren van L2 perceptuele categorieën door oude volwassenen (Ingvalson et al., 2017; Kubo & Asahane-Yamada, 2006; Maddox et al., 2013; Tajima et al., 2002), en liet zien dat er verschillen zijn in hoe jongere en oudere volwassenen leren van bepaalde aspecten van de instructie.

Conclusies

Het doel van dit proefschrift was (1) om een methodologische bijdrage te maken aan het onderzoek naar L2 spraakperceptie in natuurlijke contexten en (2) om te onderzoeken hoe verschillende leermechanismes, zowel impliciet als expliciet, bijdragen aan het verbeteren van L2 spraakperceptie.

Twee methodologische innovaties werden ontwikkeld, beschreven en gevalideerd. De eerste methodologische bijdrage is het aantonen dat het dictee (Hoofdstuk 2) gebruikt kan worden als een waardevol instrument voor onderzoek naar de perceptie van informele, continue spraak als de transcripties geanalyseerd worden met een aantal maten die gedetailleerdere informatie geven over verschillende aspecten van de input die luisteraars kunnen herkennen. De tweede methodologische bijdrage, het buiksprekersparadigma (Hoofdstuk 3), maakt het mogelijk om spraakverwerking te onderzoeken in de context van een natuurlijke dialoog waarbij zowel ecologische validiteit als experimentele controle over fonetische blootstelling worden gewaarborgd.

Drie leermechanismes voor L2 spraakperceptie werden onderzocht: impliciete lexicale begeleiding, interactieve correctieve feedback en expliciete fonetische instructie. Ten eerste werd aangetoond dat impliciete lexicale begeleiding het perceptueel leren van een nieuw L2 accent bevordert in dialoog. Dit is het eerste bewijs dat dit leermechanisme zelfs werkt in de cognitief belastende context van L2 luisteren tijdens een communicatieve taak (Hoofdstukken 4 een 5). Ten tweede werd aangetoond dat interactieve correctieve feedback leidt tot verbetering van L2 spraakperceptie waarbij expliciete feedback effectiever is dan impliciete feedback (Hoofdstuk 4). Ten derde werd aangetoond dat expliciete fonetische uitleg met minimale woordparen leidde tot verbetering van fonologisch bewustzijn en perceptie van twee L2 klankcontrasten bij zowel jongere als oudere volwassen luisteraars (Hoofdstuk 6). Dit toont aan dat het voordelig is om de aandacht van L2 luisteraars expliciet te richten op specifieke klanken en fonetische details.

Alles samengenomen laat het onderzoek in dit proefschrift zien dat het combineren van inzichten uit de onderzoeksgebieden van fonetiek, psycholinguïstiek en tweedetaalverwerving ons wetenschappelijk begrip van L2 perceptueel leren in natuurlijke contexten kan vergroten.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(1), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Babel, M., McGuire, G., Walters, S. & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006

Baese-Berk, M. (2018). Chapter one–Perceptual learning for native and non-native speech. *Psychology of Learning and Motivation*, *68*, 1–29. https://doi.org/10.1016/bs.plm.2018.08.001

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

Baese-Berk, M. M. & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36. https://doi.org/10.1016/j.jml.2015.10.008

Bates, D., Maechler, 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

Bent, T. & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Best, C. T., & Tyler, M. D. (2007). Non-native and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second-language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Bieber, R. E. & Gordon-Salant, S. (in press). Improving older adults' understanding of challenging speech: Auditory training, rapid adaptation and perceptual learning. *Hearing Research*. https://doi.org/10.1016/j.heares.2020.108054

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Booij, G. (1999). *The phonology of Dutch*. Oxford University Press.

Bowerman, S. (2008). White South African English: phonology. In R. Mesthrie (Ed.), *Varieties of English: Vol. 4. Africa, South and Southeast Asia* (pp. 164–176), Walter de Gruyter.

Bradlow, A. R., Akahana-Yamada, R., Pisoni, D. B. & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A.R. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brand, S. & Ernestus, M. (2018). Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology*, *71*, 1240–1259. https://doi.org/10.1080%2F17470218.2017.1313282

Branigan, H. P., Pickering, M. J. & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–25. https://doi.org/10.1016/j.cognition.2006.05.006

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

Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction*, *1*, 67–86. https://doi.org/10.1007/BF00158952

Broersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *The Journal of the Acoustical Society of America*, *127*(3), 1636–1644. https://doi.org/10.1121/1.3292996

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*(7-8), 1205–1224. https://doi.org/10.1080/01690965.2012.660170

Broersma, M. & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*(1), 22–34. https://doi.org/10.1016/j.system.2007.11.003

Brouwer, S., Mitterer, H. & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken work recognition. *Language and Cognitive Processes*, *4*, 539–571. https://doi.org/10.1080/01690965.2011.555268

Brown, A. (1995). Minimal pairs: Minimal importance? *ELT Journal*, *49*(2), 169–175. https://doi.org/10.1093/elt/49.2.169

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*(4), 436–458. https://doi.org/10.1177/1362168814563200

Brown, G., Anderson, A., Yule, G. & Shillcock, R. (1983). *Teaching Talk.* Cambridge, U.K.: Cambridge University Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. doi:10.3758/BRM.41.4.977

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

Burnham, D., Joeffry, S. & Rice, L. (2010). Computer- and human-directed speech before and after correction. In M. Tabain, J. Fletcher, D. Grayden, J. Hajek & A. Butcher (Eds.), *Proceedings of the 13th Australasian International Conference on Speech Science and Technology* (pp. 13–17).

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, *3*(3), 229–242. https://doi.org/10.1207/s15434311laq0303_1

Chandrasekaran, B., Yi, H., Smayda, K. E. & Maddox, W. T. (2016). Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics*, *78*, 566–582. https://doi.org/10.3758/s13414-015-0999-x

Chen, Y. & Pederson, E. (2017). Directing attention during perceptual training: A preliminary study of phonetic learning in Southern Min by Mandarin speakers. In *Proceedings of Interspeech 2017* (pp. 1770–1774). https://doi.org/10.21437/Interspeech.2017-1600

Clopper, C. G. & Bradlow, A. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*, *51*(3), 175–198. https://doi.org/10.1177/0023830908098539

CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict (Version 0.7b).

Cohn, M., Segedin, B.F. & Zellou, G. (2019). Imitating Siri: Socially-mediated vocal alignment to device and human voices. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 1813–1817). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Collins, B. & Mees, I. M. (1996). *The phonetics of English and Dutch* (3rd ed.). Leiden: E. J. Brill.

Cooper, A. & Bradlow, A. (2018). Training-induced pattern-specific phonetic adjustments by first and second language listeners. *Journal of Phonetics*, *68*, 32–49. https://doi.org/10.1016/j.wocn.2018.02.002

Cox., F. & Palethorpe, S. (2008). Reversal of short front vowel raising in Australian English. *Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2008)*, 342-345. Brisbane, Australia.

Crystal, T. H. & House, A. S. (1988). The duration of American-English vowels: An overview. *Journal of Phonetics*, *16*, 263–284. https://doi.org/10.1016/S0095-4470(19)30500-5

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A., Burchfield, L. A. & Antoniou, M. (2018). Factors affecting talker adaptation in a second language. In J. Epps, J. Wolfe, J. Smith & C. Jones (Eds.), *Proceedings of the 17th Australasian International Conference on Speech Science and Technology* (pp. 33-36). Sydney, Australia.

Drouin, J. R. & Theodore, R. M. (2018). Lexically guided perceptual learning is robust to task-based changes in listening strategy. *The Journal of the Acoustical Society of America*, *144*(2), 1089–1099. https://doi.org/10.1121/1.5047672

Drozdova, P., Van Hout, R. & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, *19*(5), 914–920. https://doi.org/10.1017/S136672891600002X

Ellis, R. (1999). Theoretical perspectives on interaction and language learning. In Ellis, R. (Ed.), *Learning a second language through interaction* (pp. 3–31). John Benjamins.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ernestus, M. & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(SI), 253–260. https://doi.org/10.1016/S0095-4470(11)00055-6

Ernestus, M., Baayen, H. & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173. https://doi.org/10.1006/brln.2001.2514

Ernestus, M., Dikmans, M. & Giezenaar, G. (2017). Advanced second language learners experience difficulties processing reduced word pronunciation variants. *Dutch Journal of Applied Linguistics*, *6*(1), 1–20. https://doi.org/10.1075/dujal.6.1.01ern

Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680. https://doi.org/10.3758/BF03212164

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

Garrod, S. & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, *15*(2), 87–105. https://www.jstor.org/stable/30029508

Giles, H. & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Mahwah, NJ: Lawrence Erlbaum.

Goo, J., Granena, G., Yilmaz, Y. & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revised and updated. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 443–482). Amsterdam: John Benjamins.

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *Journal of Rehabilitation Research & Development*, *42*(4), 9–24. https://www.doi.org/10.1682/JRRD.2005.01.0006

Guion, S. G. & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 57–77). Amsterdam: John Benjamins.

Hanulíková, A. & Ekström, J. (2017). Lexical adaptation to a novel accent in German: A comparison between German, Swedish, and Finnish listeners. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)* (pp. 1784-1788). Stockholm, Sweden.

Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, *11*(2), 186–197. https://doi.org/10.1080/15434303.2014.895829

Hillenbrand, J. M., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and Speech*, *54*(2), 241–264. https://doi.org/10.1177%2F0023830910397499

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, *33*(9), 1174–1178. https://doi.org/10.1121/1.1908941

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(2), 129–140. https://doi.org/10.1017/S0272263105050084

Irvine, P., Altai, P. & Oller Jr., J. R. (1974). Cloze, dictation, and the Test of English as a Foreign Language. *Language Learning*, *24*(2), 245–252. https://doi.org/10.1111/j.1467-1770.1974.tb00506.x

Iverson, P., Hazan, V. & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Janse, E. & Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, *39*(3), 330–343. https://doi.org/10.1016/j.wocn.2011.03.005

Janssen, C., Segers, E., McQueen, J. M. & Verhoeven, L. (2015). Lexical specificity training effects in second language learners. *Language Learning*, *65*(2), 358–389. https://doi.org/10.1111/lang.12102

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis*. Proceedings of the 1st Session of the 10th International Symposium. 29–54. Tokyo, Japan: The National International Institute for Japanese Language.

Kemps, R., Ernestus, M., Schreuder, R. & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*, 117–127. https://doi.org/10.1016/S0093-934X(03)00425-5

Kent, R. D., (Ed.). (1992). *Intelligibility in Speech Disorders: Theory, measurement, and management*. Amsterdam: John Benjamins.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kiesling, S. F. (2006). English in Australia and New Zealand. In B. B. Kachru, Y. Kachru. & C. L. Nelson (Eds.), *The Handbook of World Englishes* (pp. 74-89). Blackwell Publishing Ltd.

Kissling, E. M. (2014). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research*, *19*(3), 254–275. https://doi.org/10.1177%2F1362168814541735

Krenca, K., Segers, E., Chen, X., Shakory, S., Steele, J. & Verhoeven, L. (2020). Phonological specificity relates to phonological awareness and reading ability in English-French bilingual children. *Reading and Writing*, *33*, 267–291. https://doi.org/10.1007/s11145-019-09959-2

Kubo, R. & Akahane-Yamada, R. (2006). Influence of aging on perceptual learning of English phonetic contrasts by native speakers of Japanese. *Acoustical Science and Technology*, *27*(1), 59–61. https://doi.org/10.1250/ast.27.59

Kuhlen, A.K. & Brennan, S.E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*, 54–72. https://doi.org/10.3758/s13423-012-0341-8

Leach, L. & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

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. (2016a). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*(1), 35–64. https://doi.org/10.1017/S0272263115000194

Lee, A. H. & Lyster, R. (2016b). Effects of different types of corrective feedback on receptive skills in a second language: A speech perception training study. *Language Learning*, *66*(4), 809–833. https://doi.org/10.1111/lang.12167

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%2Fs13428-011-0146-0

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics-Doklandy*, *10*(8), 707–710.

Llurda, E. (2005). *Non-native language teachers: Perceptions, challenges, and contributions to the profession*. New York: Springer.

Long, M. H. (1980). *Input, interaction, and second language acquisition*. Unpublished doctoral dissertation, University of California, Los Angeles.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, *19*, 36–66. https://doi.org/10.1017/S0272263197001034

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*(3), 405–430. https://doi.org/10.1093/applin/ami051

Mackey, A., Gass, S. & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, *22*(4), 471–497. https://doi.org/10.1017/S0272263100004010

Maclagan, M. & Hay, J. (2007). Getting *fed* up with our *feet*: Contrast maintenance and the New Zealand English “short” front vowel shift. *Language Variation and Change*, *19*(1), 1-25. https://doi.org/10.1017/S0954394507070020

Maddox, W. T., Chandrasekaran, B., Smayda, K. & Yi, H. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, *28*(4), 1042–1056. https://doi.org/10.1037/a0034969

Major, R. C., Fitzmaurice, S. M., Bunta, F. & Balasubramanian, C. (2005). Testing the effects of regional, ethnic, and international dialects of English on listening comprehension. *Language Learning*, *55*(1), 37–69. https://doi.org/10.1111/j.0023-8333.2005.00289.x

Matthews, J. & O’Toole, J. M. (2015). Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, *28*(4), 364–382. https://doi.org/10.1080/09588221.2013.864315

Maye, J., Aslin, R. N. & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357

McQueen, J. & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception* (pp. 233–236).

McQueen, J. M., Cutler, A. & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

McQueen, J. M., Norris, D. & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech*, *49*(1), 101–112. https://doi.org/10.1177/00238309060490010601

Mitterer, H. & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS One*, *4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785

Morris, S. (1983). Dictation—a technique in need of reappraisal. *ELT Journal*, *37*(2), 121–126. https://doi.org/10.1093/elt/37.2.121

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C. & Levin, L. (2016). PanPhon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016: Technical Papers Osaka*, 3475–3484.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accent speech. *Studies in Second Language Acquisition*, *20*(2), 139–154. https://doi.org/10.1017/S0272263198002022

Neger, T. M., Rietveld, T. & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in Human Neuroscience*, *8*, 628. https://doi.org/10.3389/fnhum.2014.00628

Nerbonne, J. & Heeringa, W. (1997). Measuring dialect distance phonetically. In *Computational Phonology: 3rd Meeting of the ACL Special Interest Group in Computational Phonology*.

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

O'Malley, J.M., Chamot, A.U. & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, *10*(4), 418–437. https://doi.org/10.1093/applin/10.4.418

Oller, J. W. & Streiff, V. (1975). Dictation: A test of grammar-based expectancies. *ELT Journal*, *30*(1), 25–36. https://doi.org/10.1093/elt/XXX.1.25

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

Park, K. & Kim, J. (2018). g2p_en. https://github.com/Kyubyong/g2p (Version 1.0.0).

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, *127*(2), EL54–59. https://doi.org/10.1121/1.3292286

Peelle, J. E. & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, *42*(Suppl2), 2S11–2S16. https://doi.org/10.1044/1059-0889(2012/12-0030)

Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–226. https://doi.org/10.1017/S0140525X04220052

Pinet, M., Iverson, P. & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–1662. https://doi.org/10.1121/1.3613698

Pinet, M., Iverson, P. & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *The Journal of the Acoustical Society of America*, *130*, 1653–1662. https://doi.org/10.1121/1.3613698

Podlubny, R. G., Nearey, T. M., Kondrak, G. & Tucker, B. V. (2018). Assessing the importance of several acoustic properties to the perception of spontaneous speech. *The Journal of the Acoustical Society of America*, *143*(4), 2255–2268. https://doi.org/10.1121/1.5031123

Poellmann, K., McQueen, J. M. & Mitterer, H. (2011). The time course of perceptual learning. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the. 17th International Congress of Phonetic Sciences*, *2011, Hong Kong* (pp. 1618–1621).

Polka, L. & Bohn, O-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://psycnet.apa.org/doi/10.1016/S0167-6393(02)00105-X

Porretta, V. J., & Tucker, B. V. (2014). Perception of non-native consonant length contrast: The role of attention in phonetic processing. *Second Language Research*, *31*(2), 239–265. https://doi.org/10.1177%2F0267658314559573

Rassaei, E. (2013). Corrective feedback, learners' perceptions, and second language development. *System, 41*(2), 472–483. https://doi.org/10.1016/j.system.2013.05.002

Raveh, E., Steiner, I., Siegert, I., Gessinger, I. & Möbius, B. (2019). Comparing phonetic changes in computer-directed and human-directed speech. In P. Birkholz & S. Stone (Eds.), *Studientexte zur Sprachkommunikation Band 93: Elektronische Sprachsignalverarbeitung 2019* (pp. 42–49). Dresden: TUDpress.

Reinisch, E. & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 539–555. https://doi.org/10.1037/a0034409

Reinisch, E., Weber, A. & Mitterer, H. (2013). Listeners retune phoneme category boundaries across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75-86. https://doi.org/10.1037/a0027979

Riek, L. D. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*, 187–224. https://doi.org/10.1017/S0142716417000418

Samuel, A. G. & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/app.71.6.1207

Savignon, S. J. (1982). Dictation as a measure of communicative competence in French as a second language. *Language Learning*, *32*(1), 33–47. https://doi.org/10.1111/j.1467-1770.1982.tb00517.x

Scharenborg, O. & Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*, *75*, 525–536. https://doi.org/10.3758/s13414-013-0422-4

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schuhmann, K. S. (2014). Perceptual learning in second language learners [Doctoral dissertation, Stony Brook University]. SBU DSpace Repository. http://hdl.handle.net/11401/77746

Segedin, B.F., Cohn, M. & Zellou, G. (2019). Perceptual adaptation to device and human voices: Learning and generalization of a phonetic shift across real and voice-AI talkers. In *Proceedings of Interspeech 2019* (pp. 2310–2314). https://dx.doi.org/10.21437/Interspeech.2019-1433

Shepard, C. A., Giles, H. & Le Poire, B. A. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The New Handbook of Language and Social Psychology* (pp. 33–56). New York: Wiley.

Siegel, J. & Siegel, A. (2015). Getting to the bottom of L2 listening instruction: Making a case for bottom-up activities. *Studies in Second Language Learning and Teaching*, *5*(4), 637–662. https://doi.org/10.14746/ssllt.2015.5.4.6

Sjerps, M. J., Decuyper, C. & Meyer, A. S. (2020). Initiation of utterance planning in response to pre-recorded and "live" utterances. *Quarterly Journal of Experimental Psychology*, *73*(3), 357–374. https://doi.org/10.1177%2F1747021819881265

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America*, *101*(4), 2278–2288. https://doi.org/10.1121/1.418208

Stansfield, C. W. (1985). A history of dictation in foreign language teaching and testing. *The Modern Language Journal*, *69*(2), 121–128. https://doi.org/10.1111/j.1540-4781.1985.tb01926.x

Svalberg, A. M. L. (2007). Language awareness and language learning. *Language Teaching*, *40*(4), 287–308. https://doi.org/10.1017/S0261444807004491

Tajima, K., Akahane-Yamada, R. & Yamada, T. (2002). Perceptual learning of second-language syllable rhythm by elderly listeners. In Hansen, J. H. L. & Pellom, B. (Eds.), *Proc. 7th International Conference on Spoken Language Processing (INTERSPEECH 2002)*, pp. 249–252. Retrieved from https://www.isca-speech.org/archive/archive_papers/icslp_2002/i02_0249.pdf

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

Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. Durham, NC: Duke University Press.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203. https://doi.org/10.1017/S0272263100012870

Torreira, F., & Ernestus, M. (2010). The Nijmegen corpus of casual Spanish. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, et al. (Eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)* (pp. 2981–2985). Paris: European Language Resources Association (ELRA).

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 International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 1094–1098). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

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. doi:10.1075/ml.11.3.03tuc

Van der Feest, S. V. H. & Swingley, D. (2011). Dutch and English listeners' interpretation of vowel duration. *The Journal of the Acoustical Society of America*, *129*(3), EL57–EL63. https://doi.org/10.1121/1.3532050

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 Gasteren, N. A. M. (2019). *Improving foreign language listening through subtitles: The effects of subtitle language and proficiency on Dutch high school and university students' perceptual learning in English* [Master's thesis, Radboud University]. The Radboud University Thesis Repository. https://theses.ubn.ru.nl/handle/123456789/8247

Van Heuven, W.J.B., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*(3), 287–301. https://doi.org/10.1017/S0272263100009177

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539–552. https://doi.org/10.1016/j.system.2004.09.011

Willems, R. (Ed.) (2017). *Cognitive Neuroscience of Natural Language Use*. Cambridge, UK: Cambridge University Press.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, *23*(3), 345–368. https://doi.org/10.1017/S0272263101003023

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

Zellou, G. & Cohn, M. (2020). Top-down effect of apparent humanness on vocal alignment toward human and device interlocutors. In S. Denison, M. Mack, Y. Xu & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (pp. 3490–3496). Cognitive Science Society.

Zhang, X. & Samuel, A. G. (2014). Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 200–217. https://doi.apa.org/doi/10.1037/a0033182

Appendix A (Chapter 4)

Table A1

Code Breaker critical minimal pairs

Target Word	Phonological Competitor	
	Control, Generic CF, and Contrastive CF Conditions	Lexical Guidance Condition
bed	bid	wed
beg	big	peg
bet	bit	met
better	bitter	letter
desk	disc	deck
left	lift	theft
lesson	listen	lemon
medal	middle	pedal
mess	miss	less
pen	pin	men
red	rid	wreck
rest	wrist	test
send	sinned	lend
sense	since	fence
set	sit	pet
when	win	web

Note. CF = corrective feedback

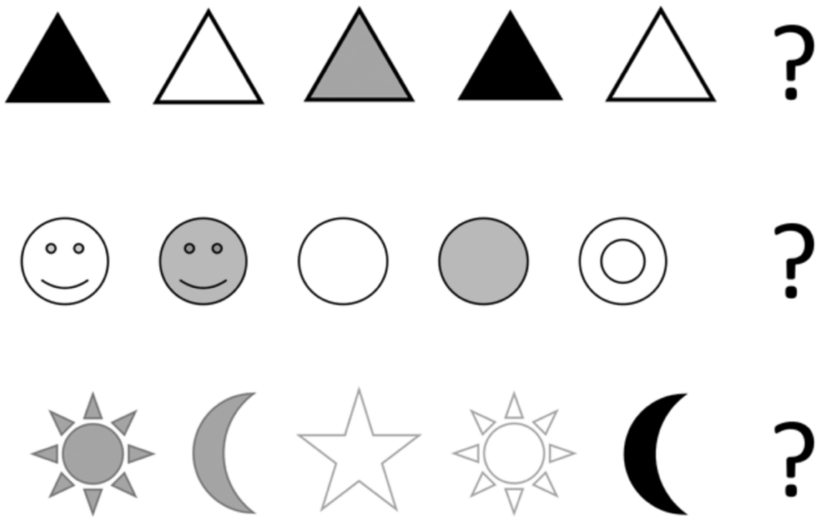

Figure A1. Three sample Code Breaker puzzles, ranging in difficulty from easier (first row) to harder (last row).

Appendix B (Chapter 4)

Table B1

Auditory lexical decision task items

Critical Words	Critical Pseudowords	Filler Words	Filler Pseudowords	Filler /ɪ/-Pseudowords
best	denner	awful	chaggard	besh
bread	fesh	busy	chobble	ched
clever	fredge	chair	choff	fredden
credit	geft	chap	cluss	frep
dress	hedden	church	cousel	jence
guest	ked	color	cupture	keff
member	lenk	couple	druse	mendow
message	lettle	crown	famper	preddle
plenty	meshin	culture	faygle	prendon
press	resk	dare	frong	spetch
ready	sester	daughter	gork	strett
spread	theck	duck	juck	zepler
		farm	kime	
		former	kire	
		funny	lasper	
		game	monder	
		gorgeous	nain	
		guide	snock	
		honest	snootle	
		huge	snop	
		judge	snurch	
		local	thosh	
		mother	trup	
		nice	tundy	
		sake		
		search		
		shine		
		south		
		square		
		struggle		
		table		
		touch		
		toy		
		upstairs		
		woman		
		youth		

Note. All “e” and “ea” vowels in stressed syllables were pronounced /ɪ/.

Appendix C (Chapter 4)

Table C1

Model predicting Code Breaker accuracy across critical trials

	β	*SE*	*z*-value	*p*-value	95% CI
(Intercept)	-4.92	1.17	-4.21	< .001*	[-7.20, -2.63]
Generic CF	-1.37	1.59	-0.86	.39	[-4.48, 1.74]
Contrastive CF	-2.00	1.58	-1.27	.21	[-5.09, 1.09]
Lexical Guidance	11.44	2.51	4.55	< .001*	[6.51, 16.37]
Trial Number	0.05	0.08	0.71	.48	[-0.10, 0.20]
Face-to-Face	-1.28	1.64	-0.78	.43	[-4.49, 1.93]
Generic CF • Trial Number	0.17	0.11	1.60	.11	[-0.04, 0.38]
Contrastive CF • Trial Number	0.34	0.11	3.12	.002*	[0.13, 0.55]
Lexical Guidance • Trial Number	0.03	0.24	0.14	.89	[-0.44, 0.51]
Generic CF • Face-to-Face	-0.67	2.47	-0.27	.79	[-5.52, 4.18]
Contrastive CF • Face-to-Face	1.96	2.22	0.88	.38	[-2.39, 6.31]
Lexical Guidance • Face-to-Face	2.12	3.23	0.65	.51	[-4.22, 8.45]
Trial Number • Face-to-Face	-0.02	0.13	-0.17	.87	[-0.27, 0.23]
Generic CF • Trial Number • Face-to-Face	-0.01	0.18	-0.06	.95	[-0.37, 0.35]
Contrastive CF • Trial Number • Face-to-Face	-0.04	0.16	-0.26	.79	[-0.36, 0.27]
Lexical Guidance • Trial Number • Face-to-Face	-0.18	0.30	-0.59	.56	[-0.76, 0.41]

Note. The condition Control and setting Computer Player are mapped onto the intercept.
CF = corrective feedback, *SE* = standard error, CI = confidence interval, * = significant.

Appendix D (Chapter 4)

Table D1

Responses to filler items in auditory lexical decision task

		Condition			
		Control	Generic Corrective Feedback	Contrastive Corrective Feedback	Lexical Guidance
Filler Words					
Acceptance Rate (%)	Mean	93.4	94.9	94.5	94.9
	(SD)	(24.9)	(22.1)	(22.9)	(22.0)
Reaction Time (ms) for “Yes” Answers	Mean	525	477	494	478
	(SD)	(251)	(237)	(234)	(229)
Filler Pseudowords					
Acceptance Rate (%)	Mean	21.1	22.6	20.8	20.6
	(SD)	(40.8)	(41.8)	(40.6)	(40.5)
Reaction Time (ms) for “No” Answers	Mean	805	763	809	777
	(SD)	(383)	(389)	(373)	(346)
Filler /ɪ/-Pseudowords					
Acceptance Rate (%)	Mean	10.3	14.5	13.0	14.5
	(SD)	(30.5)	(35.2)	(33.7)	(35.2)
Reaction Time (ms) for “No” Answers	Mean	793	809	799	823
	(SD)	(344)	(384)	(327)	(369)

We analyzed the three filler item types in Table D1 separately, using the same modeling procedures as with the critical items. For all item types, reaction times showed no significant effects of condition, but there were significant simple effects of setting: reaction times were slower in the face-to-face setting (in the Filler Words model: β = 0.09, *SE* = 0.04, *p* = .028, 95% CI [0.01, 0.17]; in the Filler Pseudowords model: β = 0.12, *SE* = 0.05, *p* = .022, 95% CI [0.02, 0.23]; in the Filler /ɪ/-Pseudowords model: β = 0.13, *SE* = 0.05, *p* = .011, 95% CI [0.03, 0.24]). Acceptance rates showed no significant effects of condition, setting, nor a condition-setting interaction for either Filler Words or Filler /ɪ/-Pseudowords. However, for Filler Pseudowords, there was one significant interaction between condition and setting: the combination of Lexical Guidance condition and face-to-face setting resulted in a lower acceptance rate for this item type, relative to the combination of Control condition and computer-player setting (β = -1.09, *SE* = 0.49, *p* = .03, 95% CI [-2.04, -0.14]).

Appendix E (Chapter 6)

Table E1

Perception stimuli

Word-final /t/-/d/ minimal pairs		/æ/-/ɛ/ minimal pairs	
beat	bead	and	end
bright	bride	bad	bed
built	build	bag	beg
cart	card	bat	bet
fate	fade	cattle	kettle
feet	feed	dad	dead
float	flowed	expanse	expense
got	god	flash	flesh
great	grade	gas	guess
greet	greed	had	head
height	hide	lag	leg
hurt	heard	land	lend
right	ride	man	men
seat	seed	mansion	mention
wrote	road	mantle	mental
sight	side	radish	reddish
slight	slide	sad	said
spent	spend	sand	send
threat	thread	than	then
white	wide	track	trek

Table E2

Phonological awareness stimuli (filler items)

Filler Minimal Pairs		Filler Homophones	
better	bitter	air	heir
bike	bake	allowed	aloud
boat	both	bare	bear
came	game	blew	blue
chase	chess	find	fined
cloud	crowd	flew	flu
desk	disc	flour	flower
file	fail	hair	hare
forgot	forget	higher	hire
fork	fort	him	hymn
fry	fly	hour	our
glue	clue	knight	night
left	lift	knot	not
lesson	listen	knows	nose
like	look	made	maid
loose	less	mind	mined
medal	middle	none	nun
note	net	peace	piece
path	bath	rays	raise
pile	pale	sail	sale
play	pray	seas	sees
pride	proud	sole	soul
rest	wrist	some	sum
rice	race	son	sun
run	pun	tale	tail
save	shave	there	their
taste	test	waist	waste
trade	train	wait	weight
true	through	way	weigh
warn	warm	wood	would

Appendix F (Chapter 6)

The scripts for each of the four phonetic instruction are provided below. The text that differs between the duration-cue and no-cue versions of the /æ/-/ɛ/ and /t/-/d/ videos is underlined.

/æ/-/ɛ/ Video with Duration Cue

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the /æ/ sound and the /ɛ/ sound. They may sound similar, but these two sounds make an important distinction in English. For example, the difference between /æ/ and /ɛ/ distinguishes words like *pan* and *pen*, and *jam* and *gem*. Do you think it's hard to hear? The /æ/ sound is usually spelled with the letter A as in *map*, while the /ɛ/ sound is usually spelled with the letter E as in *desk.*

The sounds /æ/ and /ɛ/ differ in the color, or quality, of their sound, but that's very subtle. What really helps to hear the difference is paying attention to how long the sound is.

Listen closely to these examples, in which I exaggerate the difference: *Pen. Paaan. Gem. Jaaam.* Now I'm going to pronounce the words more normally. If you listen carefully, you'll hear that the /æ/ sound is longer than the /ɛ/ sound. Try to hear the difference between *pan, pen, pan, pen, pan, pen*. Can you hear the difference in another word pair? Listen to *jam, gem, jam, gem, jam, gem*.

In short, just remember: the /æ/ sounds longer while the /ɛ/ sounds shorter. I hope that helps you!

/æ/-/ɛ/ Video with No Cue

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the /æ/ sound and the /ɛ/ sound. They may sound similar, but these two sounds make an important distinction in English. For example, the difference between /æ/ and /ɛ/ distinguishes words like "pan" and "pen", and "jam" and "gem." Do you think it's hard to hear? The /æ/ sound is usually spelled with the letter A as in "map", while the /ɛ/ sound is usually spelled with the letter E as in "desk."

The sounds /æ/ and /ɛ/ differ in the color, or quality, of their sound, but that's very subtle. Native speakers can hear the difference between the /æ/ sound and the /ɛ/ sound very easily, but for people who speak English as a second language, it can be difficult.

Listen closely to these examples, in which I exaggerate the difference: *Pen. Pan. Gem. Jam.* Now I'm going to pronounce the words more normally. Try to hear the difference between *pan, pen, pan, pen, pan, pen*. Can you hear the difference in another word pair? Listen to *jam, gem, jam, gem, jam, gem*.

It will become easier to hear the difference between /æ/ and /ɛ/ the more you practice listening. I hope that helps you!

/t/-/d/ Video with Duration Cue

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the /t/ sound and /d/ sound at the end of a word. They may sound similar, but these two sounds make an important distinction at the end of a word in English. For example, the difference between /t/ and /d/ distinguishes words like *not* and *nod*, and *bit* and *bid*. Do you think it's hard to hear? If a word ends with T or T-E , the sound is always /t/ as in *sit*. If a word ends with D or D-E, the sound is nearly always /d/ as in *did*.

The /t/ sound comes with a little puff of air, while the /d/ sound does not, but that's very subtle. What really helps to hear the difference is paying attention to how long the vowel before it sounds.

Listen closely to these examples, in which I exaggerate the difference: *Not. Noood. Bit. Biiiiid*. Now I'm going to pronounce the words more normally. If you listen carefully, you'll hear that the vowel before the /d/ sound is longer than the vowel before the /t/ sound. Try to hear the difference between *bit, bid, bit, bid, bit, bid*. Can you hear the difference in another word pair? Listen to *not, nod, not, nod, not, nod*.

In short, just remember: if the vowel is longer, you're usually hearing a /d/; if the vowel is shorter, you're usually hearing a /t/. I hope that helps you!

/t/-/d/ Video with No Cue

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the /t/ sound and /d/ sound at the end of a word. They may sound similar, but these two sounds make an important distinction at the end of a word in English. For example, the difference between /t/ and /d/ distinguishes words like *not* and *nod*, and *bit* and *bid*. Do you think it's hard to hear? If a word ends with T or T-E , the sound is always /t/ as in *sit*. If a word ends with D or D-E, the sound is nearly always /d/ as in *did*.

The /t/ sound comes with a little puff of air, while the /d/ sound does not, but that's very subtle. Native speakers can hear the difference between the /t/ sound and the /d/ sound at the end of a word very easily, but for people who speak English as a second language, it can be difficult.

Listen closely to these examples, in which I exaggerate the difference: *Not. Nod. Bit. Bid*. Now I'm going to pronounce the words more normally. Try to hear the difference between *bit, bid, bit, bid, bit, bid*. Can you hear the difference in another word pair? Listen to *not, nod, not, nod, not, nod*.

It will become easier to hear the difference between /t/ and /d/ at the end of a word, the more you practice listening. I hope that helps you!

Appendix G (Chapter 6)

Table G1

ANOVA for awareness data (full model)

	χ^2	*p*-value
Item Contrast	21.84	< .001*
Test Time	614.03	< .001*
Video Contrast	0.01	.91
Cue Information	0.02	.88
Age Group	0.08	.78
Test Time • Item Contrast	2.70	.10
Video Contrast • Item Contrast	150.22	< .001*
Test Time • Video Contrast	1.49	.22
Item Contrast • Cue Information	3.24	.07
Test Time • Cue Information	14.27	< .001*
Video Contrast • Cue Information	0.07	.79
Item Contrast • Age Group	17.06	< .001*
Test Time • Age Group	0.05	.83
Video Contrast • Age Group	0.04	.83
Cue Information • Age Group	4.05	.04*
Item Contrast • Test Time • Video Contrast	118.12	< .001*
Item Contrast • Test Time • Cue Information	21.94	< .001*
Item Contrast • Video Contrast • Cue Information	5.61	.02*
Test Time • Video Contrast • Cue Information	7.10	.01*
Item Contrast • Test Time • Age Group	0.00	.97
Item Contrast • Video Contrast • Age Group	1.09	.30
Test Time • Video Contrast • Age Group	33.63	< .001*
Item Contrast • Cue Information • Age Group	9.53	.002*
Test Time • Cue Information • Age Group	9.10	.003*
Video Contrast • Cue Information • Age Group	0.22	.64
Item Contrast • Test Time • Video Contrast • Cue Information	3.09	.08
Item Contrast • Test Time • Video Contrast • Age Group	11.11	.001*
Item Contrast • Test Time • Cue Information • Age Group	3.21	.07
Item Contrast • Video Contrast • Cue Information • Age Group	4.08	.04*
Test Time • Video Contrast • Cue Information • Age Group	5.65	.02*
Item Contrast • Test Time • Video Contrast • Cue Information • Age Group	0.26	.61

Note. * = significant.

Table G2

Separate models predicting younger and older adults' awareness accuracy

	Younger Adults		Older Adults	
	χ^2	*p*-value	χ^2	*p*-value
Test Time	334.13	< .001*	218.39	< .001*
Video Contrast	0.01	.94	0.07	.78
Cue Information	2.01	.16	1.93	.16
Item Contrast	11.09	.001*	29.36	< .001*
Test Time • Video Contrast	9.49	.002*	24.16	< .001*
Test Time • Cue Information	24.84	< .001*	0.09	.77
Test Time • Item Contrast	1.77	.18	1.00	.31
Video Contrast • Cue Information	0.01	.90	0.31	.58
Video Contrast • Item Contrast	93.70	< .001*	58.79	< .001*
Cue Information • Item Contrast	11.40	.001*	1.01	.31
Test Time • Video Contrast • Cue Information	0.05	.83	12.90	< .001*
Test Time • Video Contrast • Item Contrast	31.89	< .001*	97.78	< .001*
Test Time • Cue Information • Item Contrast	4.56	.03*	20.59	< .001*
Video Contrast • Cue Information • Item Contrast	9.52	.002*	0.03	.87
Test Time • Video Contrast • Cue Information • Item Contrast	2.68	.10	0.69	.41

Note. * = significant.

Appendix H (Chapter 6)

Table H1

ANOVA for perception data (full model)

	χ^2	*p*-value
Item Contrast	30.38	<.001*
Test Time	38.52	<.001*
Video Contrast	0.01	.94
Cue Information	0.52	.47
Age Group	40.90	<.001*
Test Time • Item Contrast	1.35	.24
Video Contrast • Item Contrast	14.83	<.001*
Test Time • Video Contrast	0.57	.45
Item Contrast • Cue Information	3.37	.07
Test Time • Cue Information	0.76	.38
Video Contrast • Cue Information	0.08	.78
Item Contrast • Age Group	0.01	.93
Test Time • Age Group	3.62	.06
Video Contrast • Age Group	0.66	.42
Cue Information • Age Group	3.68	.06
Item Contrast • Test Time • Video Contrast	43.33	<.001*
Item Contrast • Test Time • Cue Information	1.28	.26
Item Contrast • Video Contrast • Cue Information	0.17	.68
Test Time • Video Contrast • Cue Information	0.61	.43
Item Contrast • Test Time • Age Group	2.96	.09
Item Contrast • Video Contrast • Age Group	0.72	.40
Test Time • Video Contrast • Age Group	0.00	.96
Item Contrast • Cue Information • Age Group	2.59	.11
Test Time • Cue Information • Age Group	0.13	.72
Video Contrast • Cue Information • Age Group	0.29	.59
Item Contrast • Test Time • Video Contrast • Cue Information	2.09	.15
Item Contrast • Test Time • Video Contrast • Age Group	0.34	.56
Item Contrast • Test Time • Cue Information • Age Group	0.06	.81
Item Contrast • Video Contrast • Cue Information • Age Group	0.53	.47
Test Time • Video Contrast • Cue Information • Age Group	1.18	.28
Item Contrast • Test Time • Video Contrast • Cue Information • Age Group	4.21	.04*

Note. * = significant.

Table H2

Separate models predicting young and older adults' perception accuracy

	Younger Adults		Older Adults	
	χ^2	*p*-value	χ^2	*p*-value
Test Time	29.57	<.001*	12.94	<.001*
Item Contrast	25.64	<.001*	26.58	<.001*
Cue Information	3.37	.07	0.75	.39
Video Contrast	0.44	.51	0.34	.56
Test Time • Video Contrast	0.47	.50	0.16	.69
Test Time • Cue Information	0.05	.82	0.91	.34
Test Time • Item Contrast	5.22	.02*	0.14	.70
Video Contrast • Cue Information	0.33	.56	0.04	.85
Video Contrast • Item Contrast	2.74	.10	13.37	<.001*
Cue Information • Item Contrast	6.17	.01*	0.09	.77
Test Time • Video Contrast • Cue Information	0.08	.77	1.60	.21
Test Time • Video Contrast • Item Contrast	13.93	<.001*	30.29	<.001*
Test Time • Cue Information • Item Contrast	0.95	.33	0.43	.51
Video Contrast • Cue Information • Item Contrast	0.79	.37	0.03	.85
Test Time • Video Contrast • Cue Information • Item Contrast	0.44	.51	5.93	.01*

Note. * = significant.

Acknowledgements

This dissertation is the culmination of many hard but enjoyable years of work and study, during which I was lucky to have been connected to all kinds of good people who helped me along the way and who deserve my gratitude.

First and foremost, I would like to thank my supervisors, Mirjam Broersma and Mirjam Ernestus, for their dedicated support and guidance throughout my PhD trajectory. Mirjam Broersma, you were a role model in many ways, and your positive attitude and encouragement helped carry me through some of the more difficult phases of doing research. You also deserve much credit for obtaining a Vidi grant from the Dutch Research Council (NWO) for your project proposal that led to the funding of my position and that inspired many of the ideas in this dissertation. Mirjam Ernestus, you provided valuable strategic advice about the research project as a whole while also scrutinizing every detail of my written drafts, especially when it came to statistics. Despite your busy schedule, you provided fast feedback and helped me to meet my deadlines. Thank you both for your enthusiasm for the project and for giving me the opportunity to work with and learn from you.

My paranymphs, Aurora Troncoso-Ruiz and Katherine Marcoux, also deserve a special thanks. Aurora, you were such a fun office mate throughout the years, always open for a friendly chat and eager to offer me your attention, anecdotes, advice, armchair analyses, and /æ/-/ɛ/ acoustics expertise. Moreover, it was a pleasure to work together with you in developing the ventriloquist paradigm and a relief to have your help with troubleshooting its complicated audio equipment setup, which you luckily understood better than I did. Katherine, who would have thought I would get to meet and make friends with another frugal, petite, vegetarian American bespectacled booklover? You have such a knack for bringing people together, including pulling introverts like me out of their office for a cup of coffee every now and then and encouraging us to keep in shape by climbing up eight flights of stairs after lunch every day. Thank you for everything you have done to create a positive social atmosphere both around the department and within our research group.

The coronavirus-related work-from-home policy during the last year of my PhD has made me even more grateful for what I used to take for granted, the easy access to a large group of friendly colleagues on the eighth and ninth floors of the Erasmus Building and in the Max Planck Institute. Claire and Ferdy were the ones who invited me to a coffee break chat on my very first day of work and instantly made me feel welcome. Over the years, I also enjoyed taking breaks with Annika, Aurélia, Chantal, Chen, Elly, Figen, Gert-Jan, Hannah, Hanno, Joe, Lisa, Lotte, Mónica, Polina, Robert, Saskia, Tashi, Theresa, Thijs, Tim, Wei, and Xiaoru, among many others. In addition to Aurora, I shared an office on a part-time basis with Lidy, Martijn, and Toni, who enlightened me on subjects ranging from European multi-day walking tours and the challenges of first-time home buying to human measurement techniques for speech and language pathology. Thanks to you three for brightening my day whenever you were in.

The Centre for Language Studies provided a stimulating intellectual environment for carrying out my doctoral research. Thank you to the members of the Speech Production and Comprehension group for helping to broaden my knowledge of research topics adjacent to my own and for giving me the chance to practice presenting my works-in-progress in front of a critical but fundamentally supportive crowd. In particular, thanks to Esther for your always insightful feedback about my work and for your collaboration on my study with older adult listeners, and thanks to Louis for sharing your expertise about both the theoretical and practical issues involved in using mixed effects models.

The experiments in this dissertation would not have been possible without the people and facilities of the CLS Lab. Margret van Beuningen, thank you for providing such a well-organized lab environment, for arranging the useful and interesting lab lunch meetings, and for informally encouraging me and other international students to use our Dutch with you. Bob Rosbag, thank you for your invaluable technical support that made the ventriloquist paradigm possible. Much thanks also goes to the over three hundred participants who participated in the experiments reported here and to the thesis students and student research assistants who helped test them: Elisabeth, Ellen, Emma, Kyra, Lillian, Maaike, Marjolein, Nikki, Sjoerd, and Tahnee.

As part of my PhD training, I was lucky to have many educational opportunities afforded by the Graduate School for the Humanities and the International Max Planck Research School for Language Sciences. In particular, I am grateful for the informative GSH theme lunches and for the high-quality courses offered by the IMPRS on a wide range of experimental and statistical methods as well as on personal and professional development. Special thanks to Kevin Lam for being such a friendly and approachable graduate school coordinator and for providing me and my fellow IMPRS students with individual attention throughout our PhD trajectories. Moreover, I am thankful to have had excellent teachers at Radboud in'to Languages who provided a kickstart and solid foundation for my Dutch language learning.

I might never have known that the field of second language acquisition existed were it not for Kate Coughlin sending me an unusually detailed and enthusiastic debriefing e-mail after I participated in her eye-tracking study about L2 French learners' use of prosodic cues in word segmentation during my freshman year at the University of Illinois. I am grateful for Kate's mentorship back then and for her being the first one to ever tell me about the Max Planck Institute for Psycholinguistics in the Netherlands. She also introduced me to Annie Tremblay, whom I would like to thank for supervising my undergraduate research experience with second language processing and thereby giving me more confidence to pursue my interests in this direction. Several years after that, Annie also provided me with the connection to Mirjam Broersma's open PhD position, which turned out to be a fortuitous match. Many thanks also to Nivja de Jong, who recommended me for the PhD position after having supervised my Master's internship on second language speech in Utrecht.

Two close friends deserve my thanks for their close personal support during my PhD years. When Yan and I first met each other in a generative syntax course at the Utrecht Summer School before beginning our Research Master's program, we did not yet know that our paths would stay aligned for so long as we graduated from Utrecht together and both went on to pursue PhDs in linguistics at different Dutch universities. Yan, thank you for your kindness and dedication to our friendship, for commiserating about the challenges of being a non-European citizen in the Netherlands, for cooking me all kinds of delicious Chinese dishes, for inspiring me with your work ethic, for celebrating my successes, for supporting me through my setbacks, and for sharing your wisdom. My other close friend, Anna-Sophie, entered my life as a highly enthusiastic and skillful conversation partner in our Dutch class but quickly became much more to me during the years that we worked together on our PhDs in Nijmegen. Anna-Sophie, thank you for being such a generous and empathetic friend, for encouraging me to sing with you in an intimidating but beautiful choir, for livening up my work days with lunch chats and strolls around campus together, for filling my weekends with relaxing distractions, for filling my tummy with your exquisite baked goods, and for being the best listener one could ask for.

My parents, Joy and Fred, fostered my affinity for languages from a young age. Thank you both for supporting my education as well as my adventures abroad over the years. Thanks also to my siblings, Melissa and Tom, for giving your input on various ideas related to the research in this dissertation.

My parents-in-law, Claudine and Jan, and my siblings-in-law, Martine and David, have given me an abundance of interactive second language listening experience in a wide range of natural settings. Despite providing limited corrective feedback and almost no explicit instruction, you contributed enormously to my developing understanding of Dutch. Thank you all for expressing your confidence in me and for being such a fun, enthusiastic, and welcoming second family.

Finally, a deep and heartfelt thank-you goes to Rémi, my very loving partner and self-proclaimed fellow linguist ("It's not a protected title!"). Thank you for taking the significant step of moving to Nijmegen so that we could live closer to my workplace, for letting me take the nicer office when we had to start working from home, for teaching me best practices for programming and data wrangling, for patiently enduring my occasional bouts of work-related stress, for making life outside of work as joyful as I could have dreamed, for encouraging me to pursue plenty of side projects, and for being an endless source of delight and inspiration. I can't wait to see where our lives will go from here.

Curriculum Vitae

Emily Felker was born in Peoria, Illinois in the United States in 1991. She obtained her Bachelor of Science (summa cum laude) in French and Psychology at the University of Illinois at Urbana-Champaign in 2013. After a year of teaching English to middle schoolers in Strasbourg, France, she then moved to the Netherlands and obtained her Research Master in Linguistics (cum laude) from Utrecht University in 2016. She carried out her PhD research at the Centre for Language Studies at Radboud University in Nijmegen between 2016 and 2020, as a part of both the Graduate School of Humanities and the International Max Planck Research School for Language Sciences. Since 2019, she has also been working as a teacher in the Language and Communication Department at Radboud University.

List of publications

Publications

Felker, E., Ernestus, M. & Broersma, B. (2019). Evaluating dictation task measures for the study of speech perception. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th ICPhS, Melbourne, Australia 2019* (pp. 383–387). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Felker, E., Ernestus, M. & Broersma, M. (2019). Lexically guided perceptual learning of a vowel shift in an interactive L2 listening context. In *Proceedings of the 20th Interspeech* (pp. 3123–3127).

Felker, E., Klockmann, H. & De Jong, N. (2019). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, *40*(1), 111-136.

Felker, E., Tremblay, A. & Golato, P. (2015). Traitement de l'accord dans la parole continue chez les apprenants anglophones tardifs du français. *Arborescences: Acquisition du français L2*, *5*, 28-62.

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.

Submitted manuscripts

Felker, E., Broersma, M. & Ernestus, M. (submitted). The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue.

Felker, E., Janse, E., Ernestus, M. & Broersma, M. (submitted). How explicit instruction improves phonological awareness and perception of L2 sound contrasts in younger and older adults.