

# Individual Differences in the Acquisition of a Complex L2 Phonology: A Training Study

**Adriana Hanulíková**

Basque Center on Cognition, Brain, and Language and Max Planck Institute for Psycholinguistics

**Dan Dediu**

Max Planck Institute for Psycholinguistics and Donders Institute for Brain, Cognition and Behaviour, Radboud University

**Zhou Fang**

Donders Institute for Brain, Cognition and Behaviour, Radboud University

**Jana Bašňaková**

Institute of Experimental Psychology, Slovak Academy of Sciences and Max Planck Institute for Psycholinguistics

**Falk Huettig**

Max Planck Institute for Psycholinguistics and Donders Institute for Brain, Cognition and Behaviour, Radboud University

Many learners of a foreign language (L2) struggle to correctly pronounce newly learned speech sounds, yet many others achieve this with apparent ease. Here we explored how a training study of learning complex consonant clusters at the very onset of L2 acquisition can inform us about L2 learning in general and individual differences in particular. To

---

We wish to thank Peter de Jong, Elise de Bree, and Femke Scheltinga for their invaluable help in providing the original Dutch nonword repetition items, and Alexandra Dima for discussions and suggestions on statistical analyses. We thank two anonymous reviewers for their helpful and constructive comments on a previous version of this paper. Adriana Hanulíková was supported by the European Commission under the Marie Curie Action: Intra-European fellowships for career development.

Correspondence concerning this article should be addressed to Adriana Hanulíková, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 DONOSTIA, Spain. Internet: [ahanulik@googlemail.com](mailto:ahanulik@googlemail.com)

this end, adult Dutch native speakers were trained on Slovak words with complex consonant clusters (e.g., *pstruh* /pstrux/ “trout”, *štvrť* /ʃtvrɛ/ “quarter”) using auditory and orthographic input. In the same session following training, participants were tested on a battery of L2 perception and production tasks. The battery of L2 tests was repeated twice more with 1 week between sessions. In the first session, an additional battery of control tests was used to test participants’ native language (L1) skills. Overall, in line with some previous research, participants showed only weak learning effects across the L2 perception tasks. However, there were considerable individual differences across all L2 tasks, which remained stable across sessions. Only two participants showed overall high L2 production performance that fell within 2 standard deviations of the mean ratings obtained for an L1 speaker. The mispronunciation detection task was the only perception task which significantly predicted production performance in the final session. We conclude by discussing several recommendations for future L2 learning studies.

## Introduction

Mastering the phonology of a foreign language (L2)<sup>1</sup> is particularly difficult in adulthood and is often accompanied by the persistence of a foreign accent (for reviews, see Flege, Yeni-Komshian, & Liu, 1999; Piske, MacKay, & Flege, 2001; Leather & James, 1996). Not only is the production of L2 affected, its perception often remains non-native too: adult learners hear the L2 through ears trained on their native (L1) language, resulting in difficulties in perceiving L2 sounds correctly (for an overview, see Strange, 1995). However, while many adult learners struggle to correctly pronounce and perceive new speech sounds, some achieve this with apparent ease as shown by reports of the top 10 to 20% of late learners who sound (almost) accentless in their L2 (e.g., Abu-rabia & Kehat, 2004; Birdsong, 2007; Bongaerts, 1999). It is less clear, however, whether such successful learners can be identified after only a very short exposure to an unfamiliar foreign language. We aimed to explore how a training study on the production and perception of complex consonant sequences at the very onset of L2 acquisition can inform us about L2 phonological learning and individual differences.

Previous research has often focused on pronunciation proficiency in highly experienced learners and identified several factors that contribute to the variation in success in the acquisition of L2 speech such as motivation, quality of training, amount of L2 input, socio-psychological factors, personality, general intelligence, and age of acquisition (e.g., Birdsong, 2006; Ellis, 1994; Ellis, 2004; Gardner, 1985; Guiora, 1990; Leather & James, 1996; Silverberg & Samuel, 2004). The approach of looking at the ultimate attainment rather

than the onset of L2 acquisition has the disadvantage of being strongly influenced by these diverse linguistic and nonlinguistic factors. However, even at the very onset of L2 speech acquisition, when several factors are controlled, large individual differences are often observed. The causes and consequences of these individual differences are still unclear (see Ellis, 2004, for a review). The present study investigates individual differences during the initial stage of L2 phonological acquisition and, specifically, how performance in L2 production relates to L2 perception and to L1 abilities. By training participants on complex words from an unfamiliar language, we examine the relative success in perceiving and producing these words over the course of 3 weeks. An overview of earlier research on L2 speech training studies and on the role of aptitude and L1 skills in L2 acquisition is presented, followed by the aims and hypotheses of this study.

### **Training Non-Native Speech Sounds**

Many L2 speech training studies have focused on the acquisition of L2 phoneme contrasts and have shown that while some adult learners are good at distinguishing non-native contrasts, others perform less successfully, even when training and feedback are provided (e.g., Bradlow, 2008; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Strange & Dittman, 1984; Strange, Polka, & Aguilar, 1989). In their seminal study, Strange and Dittman (1984) used synthesized stimuli to train Japanese participants to discriminate between /ɪ/ and /l/ over the course of 3 weeks. While improvement in the discrimination of these synthetic stimuli was observed for seven out of eight participants, no learning occurred in the identification of naturally produced words, suggesting no generalization from synthetic to natural stimuli.

Subsequent research has shown that improvement in learning over sessions varies not only across individuals but also as a function of the training paradigm and materials used. For example, training with only F3 onset frequency enhanced /ɪ-l/ identification in some listeners, though only a subset of these listeners showed some generalization skills beyond the trained synthetic context (Ingvalson, Holt, & McClelland, 2012). When naturally recorded English words produced by several speakers were employed in training, small but significant improvement in the /ɪ-l/ identification accuracy over the course of 3 weeks was observed, although participants' performance also depended on the phonetic context in which the phonemes occurred (Logan, Lively, & Pisoni, 1991). Based on such results, it has been suggested that training with highly variable stimuli is more likely to support processing novel, untrained stimuli (e.g., Bradlow, 2008; Logan et al., 1991; Magnuson et al., 1995). It is likely that

in a high-variability training paradigm some learners develop imprecise but functional L2 sound representations, which aid them during L2 word recognition (Bradlow, 2008). In contrast, learners' representations for the learned words in a low-variability paradigm might be precise (Strange & Dittman, 1984), but the ability to recognise words, and specifically the generalization of this ability might suffer.

Taken together, training studies show that successful L2 phoneme contrast learning and large individual differences can be observed under laboratory conditions. Several studies have adopted this training paradigm to show learning patterns across a range of speech sounds that involve rapid spectral changes, but there are very few studies looking at how training helps the perception and recognition of complex articulatory coordination such as consonant clusters. Previous studies concerned with the acquisition of consonant clusters have mainly focused on universal and language specific effects on production errors such as deletion, epenthesis, or substitution etc. (see e.g., Davidson, 2010, 2011; Hanulíková & Dietrich, 2008) and on the link between perception and production (e.g., Altenberg, 2005; Kabak & Idsardi, 2007; Lee & Cho, 2005; Shibuya, 2005; Shin & Iverson, 2011). More recently, Davidson (2010) examined the effect of different input modalities and suggests that L2 speakers are generally more accurate in producing unfamiliar (Russian) consonant clusters when orthography along with audio is presented as compared to audio presentation only. Studies employing a training paradigm to investigate individual differences in how the perception of articulatory coordination relates to and improves production are lacking. Moreover, most previous studies did not focus on the characteristics of those adults who succeeded in the tasks, and the role of aptitude in these studies has seldom been considered.

### **The Link Between L2 Perception and L2 Production**

Training studies on L2 phonemic contrasts have also been carried out in order to capture how and under which conditions perceptual training transfers to improvements in production. Many of these studies have focused on Japanese speakers' difficulties with the /ɹ-l/ contrast (e.g., Bradlow et al., 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). For example, Bradlow et al. (1997) showed that improved identification of the contrast after training resulted in an improvement of /ɹ-l/ productions, as judged by native speakers of English comparing posttest with pretest performance. Bradlow et al. (1999) showed that perceptual training procedures result in long-term modifications in both L2 phoneme contrast perception and production for most learners. Similar results were obtained for the acquisition of suprasegmentals such as

Mandarin tone (see Wang, Jongman, & Sereno, 2003). Such results confirm claims that adult learners do not lose their learning capacity to perceive and subsequently produce L2 speech sounds (Flege, 1995), despite the presence of large individual differences (e.g., Bradlow et al., 1997; Flege, MacKay, & Meador, 1999).

One important question that has emerged from these perception-production training studies concerns the relationship between perception and production skills in L2. Several studies suggest that L2 perception and production are linked but the relationship is complex, even more so with increasing L2 knowledge (see Llisterri, 1995, for a review). Overall, the results of previous studies do not offer a clear picture, with several showing no correlation between the participants' ability to produce and to perceive a given contrast, a segment, or a consonant sequence (e.g., Darcy, Park, & Yang, 2011a; de Jong, Hao, & Park, 2009; Golestani & Pallier, 2007; Kabak & Idsardi, 2007; Sheldon & Strange, 1982; Shin & Iverson, 2011), while others have reported correlations between them (e.g., Rochet, 1995; Flege, 1993, 1995; Flege, Bohn, & Jang, 1997). For example, Flege (1993, 1995) claims that L2 production accuracy is necessarily limited by perceptual accuracy and that, especially in early stages of L2 acquisition, L2 production might be less native-like than perception. On the other hand, Flege failed to find a correlation between the perception and production performance of individual participants. In a further study, Flege et al. (1997) found significant but rather weak correlations between the perception and production of English vowels.

Only a few studies have examined how perceptual training leads to changes in both production and perception over time (but see Bradlow et al., 1997; Bradlow et al., 1999). For example, Bradlow et al. (1997) demonstrated that overall improvement in production occurs even without the inclusion of overt production in the training phase, but there was no correlation between perceptual learning and production performance within individual participants. Bent (2005) obtained similar results for the perception and production of Mandarin tones by naïve English speakers (but see Wang et al., 2003, who found a correlation for experienced learners).

Most previous studies focus on the acquisition of L2 segmental contrasts or suprasegmental features such as tone, leaving individual differences in the acquisition of articulatory coordination of segments underrepresented. The ability to correctly process complex consonant sequences requires good segmentation skills and thus clearly differs from the ability to perceive single phoneme contrasts or tones. For example, a word such as *prst* “finger” requires good phonological segmentation skills as well as good phonological short-term

memory resources in order to correctly perceive and reproduce the articulatory coordinations. If learners have insufficient skills, they will be likely to produce a simplified/reduced version of the word such as *pst*, *pet* etc. In the case of a single phoneme contrast, which previous studies have often presented in a simple syllable context, no such segmentation skills are required. Thus in order to perceive and produce the contrast between syllables such as *ra* and *la*, learners have to be able to process rapid spectral changes and tune into the fine phonetic details of the contrast. Tones on the other hand are manifested in changes of the fundamental frequency (F0), that is, changes in pitch spread over syllables. L2 learners of tonal languages who have no experience with tone have been shown to attend to global features of the stimuli (e.g., Bent, 2005). Taken together, it seems that there are several temporal differences between single segments, clusters and tones, which might require slightly different perceptual and sensory-motor activities and memory resources. Because some of the above studies have shown that the production system is flexible enough to learn non-native features after training, it should be possible to observe similar learning effects for complex consonant clusters. The question remains as to whether learning clusters can be observed within a short training paradigm or whether there are restrictions in terms of what late L2 learners can learn to produce and perceive, especially in the initial stages of learning novel phonotactic constraints.

### **Aptitude and L1 Skills as Predictors of L2 Learning**

The role of aptitude in L2 acquisition has been approached within the framework of developing standardized tools for L2 skills assessments (e.g., the Modern Language Aptitude Test [MLAT]; Carroll & Sapon, 1959). As a standardized measure for language aptitude, the primary aim of these projects was to predict learners' aptitude for learning a new language before starting L2 acquisition (see Robinson, 2005, for a recent review of studies on aptitude in L2 acquisition). All of the major aptitude tests suggest that one of the strongest predictor variables in L2 success is phonetic coding ability (Carroll, 1981; Sparks et al., 1997), an ability to identify distinct sounds or strings of sounds, to form associations between those sounds and symbols representing them, and to retain these associations. It is less clear, however, which factors are responsible for such phonetic coding ability. This ability could be an innate aptitude (Carroll, 1973) and, in this case, aptitude in L2 could partially be a residue of L1 aptitude. This would suggest that there might be a relationship between L1 phonetic skills and L2 phonetic skills. But aptitude could also depend on the amount of experience with other languages (McLaughlin, 1990). Thus the likelihood of

finding exceptional learners with high aptitude might be higher in groups of learners with more experience and mastery in many different languages, such as Dutch university students, who tend to be multilingual, than in monolingual groups. Thus it could be that exceptional learners at the onset of L2 acquisition are easier to find in such a population.

In addition to aptitude, some studies suggest that success in foreign language acquisition might be related to L1 skills and general cognitive abilities (Darcy, Park, & Yang, 2011b; Sparks & Ganschow, 1993; Sparks, Ganschow, & Patton, 1995; Sparks et al., 1997). For example, Sparks and Ganschow (1993) demonstrated that less successful (high-school) students of a second language might have native language learning difficulties, primarily with phonological processing. Sparks et al. (1995) found evidence that aptitude (as measured by MLAT scores) and L1 native phonological-orthographic skills (tested in tasks such as word recognition, spelling, phonemic awareness) impact on learning of a foreign language. In a further study, Sparks et al. (1997) followed up on a number of participants from their previous studies through the second year of L2 acquisition and confirmed previous findings that performance on L1 vocabulary tests (Peabody Picture Vocabulary Test) predicted L2 proficiency.

While these studies are concerned with the general success in L2 learning, there are only a few studies that focus directly on how L1 phonetic capabilities may predict successful learning of a new complex phonological system. For example, Diaz, Baus, Sescera, Costa, and Sebastián-Gallés (2008) used a mismatch-negativity design to examine individual differences in L2 phonetic acquisition and demonstrated that adult learners' L2 phoneme contrast perception correlates with their discriminatory abilities for both native and unfamiliar (Finnish) phonemes. This means that poorer perceivers of an L2 phonological contrast have poorer discrimination performance in their L1, while better perceivers have better discrimination performance.

### **Aims of the Present Study**

The aim of this study was twofold: (1) to extend the study of individual differences during acquisition of unfamiliar consonant sequences and to explore whether successful learners can be identified at the very onset of L2 acquisition and (2) to explore how performance in L2 production of consonant clusters relates to performance in various L2 perception tasks as well as to native-language performance. Thus the question is whether initial non-native language perception and production abilities relate to training-induced learning and to general cognitive abilities and native language skills. This specifically pertains to the question of what factors contribute to individual variation across learners with

similar language backgrounds when several variables (e.g., age, L2 experience, initial exposure, and training) are controlled.

To this end, we trained native speakers of Dutch to perceive and produce Slovak words with complex consonant clusters (e.g., *pstruh* /pstrux/ “trout”, *štvrt'* /ʃtvrc/ “quarter”) that do not exist in the Dutch language. In previous studies, individual variation in L2 perception or production has usually been measured by means of performance variability and the relationship between performance on various sessions including or excluding explicit training (for reviews, see Bradlow, 2008; Ellis, 2004; Llisterri, 1995). We developed a training paradigm with three testing sessions (with a short training taking place at the beginning of the first session), separated by one week. The participants were tested on a battery of tasks aimed at assessing (a) their accuracy in perceiving, recognizing and producing novel consonant sequences and (b) their L1 skills.

The accuracy in L2 receptive phonology was tested using phoneme monitoring, word monitoring, and mispronunciation detection. The accuracy in L2 productive phonology was tested using imitation and reading. All perception and production tasks were administered in each of the three sessions. The three perception tasks were chosen because we hypothesized that they are likely to tap into partially different processing levels. Phoneme monitoring can reflect how easy it is to extract sublexical information. Word monitoring is good for looking at word recognition in context, and it might reflect segmentation skills at the lexical and sublexical level. Mispronunciation detection might tap into the listeners' ability to match the orthographic representation of a whole word with the phonological representation. For the production tasks we chose reading and imitation, which pose different problems to the learner. While imitation might simply involve the ability to retain and repeat an acoustic exemplar of an auditorily presented word, reading a foreign word requires the ability to segment the word into individual segments, to assign phonological representations to segments and finally to produce them. This means that, in order to succeed in L2 reading, the learner has to be aware of the L2 speech sounds that differ from their native language. The reading task was chosen because previous research (e.g., Sparks et al., 1997) has demonstrated that foreign language word decoding is a good predictor of overall proficiency and of oral proficiency in a foreign language. However, both imitation and reading in the L2 also require certain sensorimotor flexibility for producing complex consonant sequences. This flexibility was partly assessed within the L1 control tasks using nonword repetition.

The participants' native language skills were assessed at the end of the first session.<sup>2</sup> A set of control tasks was administered to test the participants' L1

lexical fluency skills and short-term phonological working memory capacity, using semantic and phonetic fluency tasks (also called category and letter fluency tasks) as well as the nonword repetition task. The control tasks were partly chosen based on previous research (Sparks et al., 1997), where native language vocabulary skills (e.g., word recognition, spelling, phonemic awareness) predicted the overall proficiency in a foreign language. Yet other studies suggest that measures of phonological working memory (e.g., nonword repetition) predict success in foreign language acquisition (e.g., Gathercole, Service, Hitch, Adams, & Martin, 1999; Papagno & Vallar, 1995).

We hypothesized that a short auditory training session with orthographic representations of L2 complex words would suffice to induce improvement in perception and production of consonant clusters over the course of 3 weeks (in line with some previous studies on L2 phoneme contrasts). Second, we expected a large amount of individual variation, to the extent that it would be possible to identify a few highly successful learners even at this early stage. Third, we explored whether L1 skills such as nonword repetition, phonetic and semantic fluency predict L2 production (and possibly L2 perception) of consonant clusters.

## Method

### Participants

Forty native speakers of Dutch (37 females; mean age 21 years, range 18–31), all students at Radboud University, Nijmegen, received payment for their participation. Two additional participants (1 and 11) were excluded because they grew up speaking another language in addition to Dutch. All participants spoke English as a second language, and all had learned at least two other foreign languages, most often French (39 participants) and German (38 participants). In addition, knowledge of Spanish was reported by 10 participants, Latin by 7 and (Ancient) Greek by 6 participants. All participants had learned a second language in school from an average age of 11 years (range 8–13 years) and the third and fourth languages from an average age of 12 years (range 9–18 years). None of the Dutch students spoke Slovak or reported knowledge of another language where the consonant clusters that were used in the present studies are legal. The nonfamiliarity with Slovak was confirmed with a questionnaire at the end of the experiment, where participants were asked to score how familiar they were with the language they had learned in the experiment on a scale from 0 to 4 (0 = *not familiar at all*, 4 = *very familiar*); the average familiarity was 0.23. When asked to guess which language they learned in the experiment,

23 participants indicated a Slavic language, with Russian being the most frequently mentioned (9 participants) followed by Czech (7 participants). Thirteen of the students indicated that they had previously attended phonetics or phonology classes.

### Raters

For obtaining ratings of the learners' production data (see details below), we recruited a total of 74 native speakers of Slovak (36 males, ages 17–59, average age 23.8), the majority of them from the Bratislava area of the Slovak Republic. The raters were asked to listen to and evaluate the words produced by the 40 Dutch participants and one Slovak speaker (the first author, who also produced the experimental materials). They were told that they would hear non-native as well as some native speakers of Slovak. Given the large amount of words to be rated, we decided to elicit ratings only for the final session in order to capture the final achievement in pronunciation.

### Materials and Design

For the L2 perception and production tasks, we selected two sets of 36 Slovak words (nouns, verbs, and adjectives) and 36 nonwords organized in minimal pairs, all of which featured a range of complex consonant clusters (e.g., *vzblkňut'* /vzbl:knuc/ "to burst" and *\*vzblkňut'* /vzbl:kɲuc/, *štvrt'* /ʃtvrc/ "quarter" and *\*stvrť'* /stvrč/)<sup>3</sup>. The first set of words and nonwords was used in the first and second test sessions, the second set was used in the third session. For the training session, only the words but not the nonwords of the first set were used. For each word, a nonword foil was created by changing one of the phonemes within the consonant cluster (e.g., /ʃ/ to /s/, /n/ to /ɲ/; see Appendix A for a full list of words and nonwords). All consonant sequences are phonotactically illegal in Dutch (and in English, German, French and Spanish—the most frequent L2s of our participants), but only some of the phoneme distinctions are non-native to Dutch listeners (see Appendix B for further details). Two sets of 36 sentences were created using the above words and nonwords as the critical targets. Each sentence occurred in two versions, either with the correct word or with the nonword variant (e.g., *Palubné hodiny ukazujú štvrt'/\*stvrť' na desat'* /palubnɛ: fɔʒini ukazuju: ʃtvrc/\*stvrč na ʃɛsac/ "Aircraft clock shows quarter to ten"). Within each set, four additional words and sentences were used for the practice trials.

For each task, two experimental lists were created such that half of the words were correct words and the other half were nonwords. The stimuli were compiled in a random order with the constraint that a maximum of three consecutive trials were words or nonwords. Every participant was assigned to

only one list. In the first and second testing sessions, participants heard the same set of 36 words and sentences, but in the third session they heard the second set of 36 words and sentences.

The materials were read at a normal speech rate by a female native speaker of Slovak (first author). Digital recordings were made in a soundproof booth at 44.1 kHz sampling rate with 16-bit resolution. The recordings were cut into single speech files using the Praat speech editor (Boersma, 2001). All files were normalized so that their mean amplitude was approximately equal.

### **Procedure**

Participants were tested individually in a quiet room and were seated in front of a computer monitor. Auditory stimuli were played over Sennheiser headphones, the verbal responses were recorded with a high-quality Sennheiser microphone. Participants received written instructions in Dutch explaining that they were taking part in a study looking at how new words in a foreign language are acquired, and that the study would consist of several tasks administered in three sessions one week apart. For each task and in each session, participants received the following instructions. In the training phase (which took place once at the onset of the first session and lasted approx. 15 minutes), they were told that they would first see a word on the monitor (for 4 seconds) and then hear the same word. Their task on each trial was to repeat the word as correctly as possible twice in a row. Each word appeared on three successive trials. After a break, the procedure was repeated, so that altogether, participants read, heard, and produced 12 versions of each token.

In all three test sessions, all three perception tasks were administered. Phoneme monitoring consisted of the visual presentation of a target Slovak phoneme to be monitored for in an auditorily presented Slovak word or nonword, which contained the target phoneme on 50% of the trials. Participants were required to press a button if the item contained the target phoneme. In the word-monitoring task, participants were required to press a button if a visually presented Slovak word occurred in a spoken sentence, which was the case on 50% of the trials. In the mispronunciation detection task, participants saw a word and heard either a correct or an incorrect version (the nonword) of that word. They were asked to press a button if they detected an incorrect pronunciation, which was the case on 50% of the trials. The perception tasks were followed by two production tasks, reading and imitation, where participants read visually presented words and were asked to imitate spoken Slovak words.

The control tasks were administered only once at the end of the first session. In the phonetic fluency task, the participants were asked to produce as many

words starting with “m” as possible in 60 seconds. In the category fluency task, participants were asked to produce as many names of animals as possible in 60 seconds. During the nonword repetition task (the Dutch nonword repetition task is a computerized and adapted version for adults of the Dutch test by de Jong & van der Leij, 1999), participants were asked to repeat spoken Dutch nonwords. Their utterances were recorded for later scoring. All participants were assessed for their language history, their language habits and other skills in a language questionnaire at the end of the first session (see Appendix C for the full questionnaire).

### Coding and Scoring

Phoneme monitoring, word monitoring and mispronunciation detection were run in each of the three sessions such that each participant produced three sets of results per task. We scored these tasks using the percent of correct responses and the sensitivity index  $A'$ , a non-parametric variant of  $d'$ .  $A'$  quantifies the correct and incorrect responses while taking the possible response biases into account. It can take values between 0 and 1, with 1 representing perfect discriminability and 0.5 representing chance performance (Pallier, 2002; Fogarty, Baker, & Hudson, 2005).<sup>4</sup> Because percentage correct and  $A'$  correlated very well (for phoneme monitoring:  $r = 0.95, p < .001$ ; for word monitoring:  $r = 0.94, p < .001$ ; and for mispronunciation detection:  $r = 0.80, p < .001$ ), we used only  $A'$  in further analyses. To take response bias into account, we also computed  $B''_d$ , a non-parametric version of  $\beta$ , where 0 indicates no bias, positive numbers (the maximum value is 1.0) indicate a tendency to answer “no,” and negative numbers (the minimum value is  $-1.0$ ) represent a tendency to answer “yes” (see Pallier, 2002).

The scoring of the participants' utterances in reading and in imitation was done by Slovak native speakers (raters) using a WEB-based platform developed specifically for this project. The raters saw on average 140 random reading and imitation words produced by the 40 participants ensuring that each item produced by each participant in each task was rated by an average of 4 raters. The raters rated each item on a scale from 1 (very bad Slovak pronunciation) to 7 (excellent Slovak pronunciation). Ratings were obtained only for the final session in order to evaluate the production skills at the end of the short immersion in Slovak.

The participants' imitations of the nonwords in the nonword repetition task were transcribed by a native Dutch speaker and compared with the target form using the Levenshtein distance (Levenshtein, 1966), a measure of the differences between two strings taking into account insertions, deletions and

**Table 1** Summary of performance as measured by  $A'$  (standard deviation is provided in brackets) for each perception task across all sessions and for each session individually

	Phoneme monitoring	Word monitoring	Mispronunciation detection
All sessions	0.77 (0.10)	0.67 (0.14)	0.79 (0.08)
Session 1	0.77 (0.08)	0.66 (0.16)	0.78 (0.10)
Session 2	0.75 (0.12)	0.70 (0.14)	0.80 (0.07)
Session 3	0.80 (0.08)	0.66 (0.11)	0.79 (0.08)

substitutions. These scores were then normalized between 0 and 1, where 1 means best and 0 means worst performance. We scored the phonetic and semantic fluency tasks by counting the number of valid words produced by the participant.

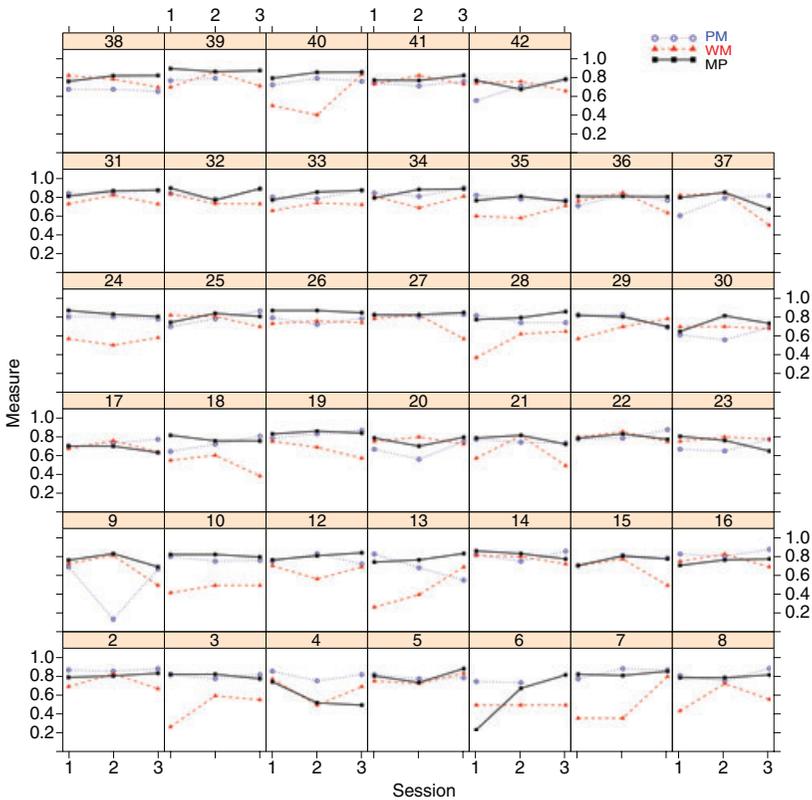
## Results

We will first present the results for all the tasks (perception, production and control) and then describe the relationship between L2 production, L2 perception, and L1 skills, ending with a brief discussion of the top learners.

### Perception tasks

Averaged across all participants and sessions, the performance in the perception tasks was clearly above chance for all three tasks. This was confirmed by  $\chi^2$  tests, comparing the frequency of response types to that expected by chance (all were highly significant at  $p < .001$ ). Table 1 shows the performance as measured by  $A'$  for all three perception tasks averaged across the three sessions as well as for each session individually. It is apparent that there were very weak learning effects (increases in  $A'$  score) across the three sessions for phoneme and word monitoring, and no learning in the mispronunciation detection.<sup>5</sup> However, as can be seen in the participants' individual performance in all three sessions in Figure 1<sup>6</sup>, there were clear learning effects for some participants and tasks.

We also tested whether participants were biased in their responses, and found that there was a strong liberal bias in the word and phoneme monitoring tasks (i.e., towards reporting the monitored phoneme or word) across all sessions, and a strong conservative bias in the mispronunciation task (i.e., towards not reporting a mispronunciation) across all sessions. These biases might indicate that the tasks were difficult so that participants probably developed various response strategies.



**Figure 1** Performance as measured by A' (1 means perfect performance, 0.5 is chance level) for the three perception tasks (PM = Phoneme Monitoring, WM = Word Monitoring, MD = Mispronunciation Detection) for each subject (one graph per subject) and each of the three sessions.

To test whether performance in the three perception tasks across the three sessions reflects a stable trait, we computed the test-retest reliabilities (correlations between sessions) for each task. The results showed moderate to high test-retest reliabilities and suggest that we are dealing with relatively stable traits. For example, the correlations between sessions 1 and 2 are  $r = 0.42$  for phoneme monitoring,  $r = 0.68$  for word monitoring, and  $r = 0.38$  for mispronunciation detection ( $p < .05$  for all).

We also tested whether there was any correlation between the three perception tasks. Significant correlations were only found in the third session, with mispronunciation detection moderately correlating with word monitoring and with a trend for a correlation with phoneme monitoring (see Table 2). The lack

**Table 2** Spearman's  $\rho$  non-parametric correlation coefficients between tasks (across participants),  $p$  values are in parentheses, bold cells show correlations significant at  $\alpha \leq 0.05$

Spearman's $\rho$	NWR	PF	SF	PM	WM	MD	Imitation
PF	0.01 (0.937)						
SF	0.16 (0.364)	<b>0.34</b> <b>(0.045)</b>					
PM	-0.09 (0.590)	-0.18 (0.270)	0.18 (0.279)				
WM	-0.01 (0.941)	0.28 (0.083)	0.23 (0.151)	0.12 (0.459)			
MD	-0.05 (0.736)	0.10 (0.519)	0.11 (0.500)	0.30 (0.068)	<b>0.36</b> <b>(0.022)</b>		
Imitation	0.03 (0.876)	0.04 (0.813)	-0.17 (0.322)	0.15 (0.357)	0.26 (0.099)	0.22 (0.178)	
Reading	0.20 (0.254)	0.20 (0.250)	0.19 (0.278)	0.26 (0.121)	0.08 (0.603)	<b>0.42</b> <b>(0.007)</b>	<b>0.51</b> <b>(0.001)</b>

*Note.* NWR = Nonword Repetition; PF = Phonetic Fluency; SF = Semantic Fluency; PM = Phoneme Monitoring; WM = Word Monitoring; MD = Mispronunciation Detection.

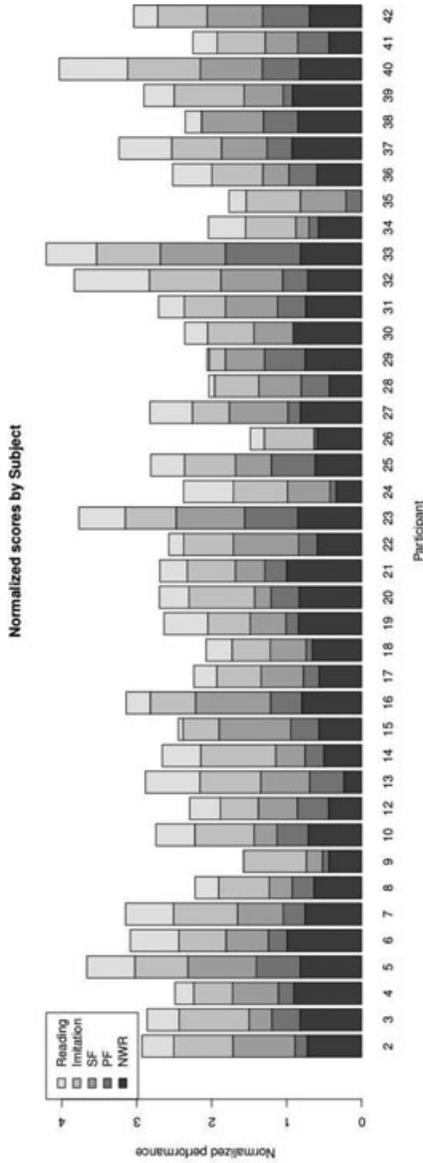
of further significant correlations suggests that the three tasks might indeed be tapping into partially different processes.

### Production tasks

The average rating (scale from 1 to 7, where 7 means native-like pronunciation) on the imitation task was 4.4 ( $SD$  0.52), and on the reading task was 3.9 ( $SD$  0.53). A native Slovak speaker (the first author) was used as a baseline comparison by presenting her utterances for rating as for any other participant, and her mean rating was 6.7 (median 7.0,  $SD$  0.99). Figure 2 shows the participants' individual rating scores for imitation and reading (normalized between 0 and 1).<sup>7</sup> It can be seen that there is substantial individual variation in both production tasks. There was a moderate positive correlation between the two production tasks across subjects ( $r = 0.51$ ,  $p = .001$ ).

### Control tasks

In the nonword repetition task, participants achieved an average score of 0.66 out of a maximum of 1 ( $SD$  0.20). In the phonetic fluency task, participants had a mean of 18 words ( $SD$  4.80, range 10–34), and in the semantic fluency



**Figure 2** Performance (normalized between 0 and 1 for presentation purposes only) for the production and control tasks (SF = Semantic Fluency, PF = Phonetic Fluency, NWR = Nonword Repetition). The horizontal axis represents the participants ordered by their numeric identifier. Each stack of bars represents the normalized performance of a participant for the respective tasks.

they had a mean of 29.25 words ( $SD$  5.49, range 16–39). Figure 2 shows the combined results of the L2 production and control tasks. As can be seen, there are substantial individual differences with a few learners showing clearly better performance across all tasks (e.g., 23, 32, 33, and 40). There was a positive correlation between phonetic and semantic fluency ( $r = 0.34, p = .045$ ).

### **Relationships Between Production, Perception and L1 Skills**

Because there was very little learning in the perception tasks across the three sessions, we assessed the relationship between perception, production and L1 skills for session three only. The results of exploratory correlational analyses are summarized in Table 2, and suggest that the only relationship between perception and production is represented by the moderate positive correlation between mispronunciation detection and reading performance ( $r = 0.42, p = .007$ ). None of the L1 control tasks predicted L2 production or perception skills.

In conclusion, mispronunciation detection appears to be the best predictor of L2 reading performance in the early stages of L2 acquisition. We will discuss below why this might be the case.

### **Successful Learners**

In order to evaluate the participants' success in L2 pronunciation at the onset of L2 phonology acquisition, we applied an index of accent-free L2 pronunciation as suggested by previous research (e.g., Flege & MacKay, 2004; Flege, 2005), whereby a non-native speaker could be considered (almost) accent-free if his/her ratings falls within two standard deviations of the mean ratings obtained by native speakers (Flege, 2005). Taking the overall L2 speakers' performance averaged over imitation and reading, two participants (participant 32 with a mean of 5.2; and participant 40 with a mean of 5.1) fell within two standard deviations of the mean ratings obtained for the native Slovak speaker. This suggests that these two participants could be considered as having approached near native-like pronunciation in Slovak. Note that this interpretation should be considered with caution due to the possibility that raters might have felt sympathetic with the L2 speakers.

Because we did not test a native speaker on the perception tasks, no such performance estimate is available for the perception success. However, we can report that the achievements in perception tasks varied considerably within participants, and two participants clearly showed best performance in all three perception tasks (32 and 34). Although we did not find any participant with top performance in all L2 and L1 tasks in all sessions (which might be hard to find even among native speakers), there were a few participants performing

overall better in most tasks than other participants. To find these participants, we computed the percentage of participants worse than a given participant in a task (the rank percentage; for example, participant 32 was better than 94.9% of the participants at imitation).<sup>8</sup> Three best learners were identified in this way (32: female, 24 years old, English language student, phonetics training, no musical skills; 33: female, 23 years old, Education sciences, phonetics training, no musical skills; and 40: female, 20 years old, Philosophy student with no phonetics training, musical skills). All three participants spoke English, German, and French as L2. Participant 32, however, also indicated that in her childhood, German and English were spoken at home, although she only started learning these languages at school age and did not report to be bilingual or to have either of these languages as her native language (therefore this participant was not excluded from the analysis). It is conceivable that this multilingual situation in early childhood contributed to the high performance of this participant.

Taken together, it is apparent that a common pattern explaining successful performance cannot be identified. However, the high performance of a few participants does fit well with previous results that L2 perception skills can be linked to L2 production skills and to L1 abilities (e.g., Sparks et al., 1997; Diaz et al., 2008) even though we did not observe evidence for such a link for all participants. It is also possible that there were other exceptional learners in our sample who need a little more time to excel at the acquisition of these complex L2 consonant clusters. Finally, to obtain a better estimate of native-like pronunciation and perception, future studies should consider more native speakers as a control group.

## Discussion and Conclusions

The study reported here is the first of its kind investigating individual differences in the training of L2 production and perception of complex consonant clusters in late second-language learners. Dutch speakers took part in a short training on unfamiliar words with various consonant clusters and were tested in three sessions on a battery of L2 perception and production tasks as well as on tasks assessing their L1 abilities. In line with some previous studies, we showed that the short training (15 minutes) and the type of training (repetition after auditory and orthographic presentation of L2 words) resulted in rather weak or no improvements in the different perception tasks. However, as expected, substantial individual variation was observed with some individuals showing strong learning effects in some tasks. Thus it is possible to identify successful

learners even at the very initial stages of L2 acquisition. It is interesting that none of the L1 skills assessed here predicted L2 perception and production performance, and that only mispronunciation detection but not the other perception tasks correlated with L2 production. We will discuss these results in turn.

Our finding that there was overall no strong learning effect across the three sessions in the perception tasks is perhaps surprising. There are some studies which have shown improvement in perception (and/or production) after training on phoneme contrasts (e.g., Bradlow et al., 1997). However, the learning effects in many studies are rather weak and seem to depend on further factors such as phonetic context (e.g., Logan et al., 1991) and thus appear to be consistent with our results. The lack of stronger learning effect could be due to the experimental set-up, using a short and low-variability training session: participants were trained on Slovak words spoken by a single speaker. It has been observed that speaker variation may help learning novel words (e.g., Bradlow, 2008; Logan et al., 1991; Magnuson et al., 1995). However, it is possible that in the case of articulatory coordination, one speaker who produces the consonant clusters in a clear way might be as good as (if not better than) having multiple speakers. With the present design, we cannot determine whether the lack of stronger learning effects is due to the low-variability training (having only one speaker), or whether participants were generally unable to form the correct phonological (and phonetic) category boundaries due to the limited amount of training. Future studies on this topic should thus test whether the inclusion of more speakers during training and/or longer training sessions with feedback enhances learning effects.

A second factor that might have precluded stronger learning effects could be the participants' motivation. This is of course not particular to the present study but a general problem for studies involving many tasks and experimental sessions. It is well known that motivation can be a powerful factor and that low motivation may cause poor performance even in learners with high aptitude (e.g., Gardner, 1985; Ellis, 2004). For future research, one possible way to select participants with more motivation to learn an unfamiliar language would be to find new learners of that respective language. The problem for these participants would not be the motivation or incentive to improve, but it could be that such a population is generally interested in languages and would show high language skills. This could then be taken as a factor in a comparison between groups of "real" learners and participants in a laboratory setting. For future work, we suggest shorter experiments with one or two sessions, which are likely to keep up the learners' motivation.

Despite the lack of strong learning effects, we found a large amount of variation between learners both in the performance level and learning trajectory. This suggests either different levels of motivation or different levels of phonetic aptitude across a participant sample with a lot of experience with learning foreign languages. It is however noteworthy that all participants succeeded to some degree in learning the rather complex and unfamiliar phonology of Slovak: none performed at chance level in the perception tasks and none had very low ratings in the production tasks. At the same time, only two of the participants excelled in L2 perception and production, and performed in their productions to a degree that was relatively close to native speaker performance (within two standard deviations from L1 speaker's mean ratings) at this early L2 stage. This could be taken as confirming previous suggestions that native-like levels of production can only (if at all) be achieved after some (perhaps substantial) training (e.g., Birdsong, 2004). However, it should be noted that 2 out of 40 participants achieved native-like levels according to Flege's (2005) criterion and thus it is conceivable that even more successful "L2 producers" can be identified with a larger sample. Having a larger sample would also increase the power to identify the differences and commonalities underlying (L2) phonetic aptitude. Future work should thus aim for very large sample sizes, ideally hundreds of participants.

Another perhaps surprising finding of the present study is that performance in the mispronunciation detection task in the final session predicted significantly performance in L2 word reading in the final session. One may ask why this task is more predictive of production abilities than other perception tasks. One possibility is that mispronunciation detection requires more substantial subvocal rehearsal of the phonological form of the word in order to be able to match the form with a stored episode of the same word and to then make a decision about the correctness of its pronunciation. An alternative explanation is that the process of evaluating the L2 speech material depends on the instructions provided in the three perception tasks. While both phoneme and word monitoring required participants to indicate whether a visually presented letter or word occurs in the speech stream, the mispronunciation detection task explicitly required judging whether a visually presented word was pronounced correctly or incorrectly. We suggest that mispronunciation detection forces listeners to parse every phoneme or combination of phonemes more carefully than phoneme and word monitoring. Future research could explore whether mispronunciation detection is indeed the perception task which is most predictive of production success of complex consonant clusters.

It is interesting that while both production tasks correlated with each other, not all perception tasks did. This indicates that these tasks might tap into distinct abilities, which could explain why any perception-production link is likely to be very complex. However, test-retest reliabilities suggest that the perception tasks (and possibly the production tasks as well) measure stable traits, amenable in principle to heritability and other types of genetic studies (see Stromswold, 2001, for a review). Such studies require a streamlining of the tasks in order to test large numbers of participants and better control for other likely contributing factors such as general intelligence (see Deary, 2012, for review).

The lack of a strong link between perception and production has been reported previously for learning of L2 phoneme contrasts (e.g., Darcy et al., 2011a; de Jong et al., 2009; Golestani & Pallier, 2007) and for articulatory coordinations (e.g., Altenberg, 2005; Kabak & Idsardi, 2007; Lee & Cho, 2005; Shibuya, 2005). The result of our study is thus in line with previous findings. It is likely that especially in the initial stages of L2 acquisition, perception and production dissociate, and learners approach the task of learning these sequences with differing strategies. It could be that only once precise perceptual representations have been established, accurate utterances can be observed. Correctly perceiving and producing such complex sequences of consonants requires good segmentation skills. The accuracy with which learners then produce such words might be limited by how accurately they segment these words into their individual parts. L2 perception accuracy may depend on the L1 phonology of the speaker too. For example, a study by Hallé, Segui, Frauenfelder, and Meunier (1998) demonstrated that French listeners misperceive nonnative consonant sequences (e.g., /dl/ was perceived as /gl/) and thus assimilate them to sound patterns that exist in their language. However, perception difficulties might not always be reflected in production difficulties or vice versa (e.g., Altenberg, 2005; Lee & Cho, 2005). This is in line with the claim that perceptual performance is not a simple inversion of phonological phenomena occurring in the production system (e.g., Kabak & Idsardi, 2007), suggesting that perception and production necessarily rely on different mechanisms and thus any link between them will be rather complex. Future research would be directed at exploring the predictive power of different perception tasks in more detail.

We observed positive (but non-significant) correlations between performance in the production tasks and nonword repetition and the fluency measures. This again suggests that future learning studies should be large-scale to increase the statistical power especially given previous evidence showing that measures of phonological working memory such as nonword repetition

predict outcomes of foreign language acquisition (e.g., Gathercole et al., 1999; Papagno & Vallar, 1995).

Finally, we believe there are two ways that may increase the interpretability of the results of training studies for L2 learning in general and individual differences in particular. One promising approach would be to focus on learning over time and have a training phase before each testing session and to increase the number of sessions (e.g., 10 sessions, each separated by a week and each, for instance, consisting of the same training, perception and production tasks). Such an approach might reveal possible interactions of phonetic aptitude with memory consolidation (Dumay & Gaskell, 2007). A second approach is to focus on the very initial stage of L2 acquisition and test performance right after the initial training at one test session only. This approach has the advantage of reducing noise caused by differing levels of motivation, alertness, fatigue, etc. and minimizes the influence of mediating factors such as (long-term) memory. It may thus be more likely to capture “pure” phonetic aptitude. On the other hand it could be argued that minimizing the influence of long-term memory removes a key component of naturalistic L2 acquisition. Moreover, it has been suggested that an important factor contributing to successful L2 sound acquisition might be the focus on perceiving full sentences and larger discourse units rather than isolated speech sounds and words (e.g., Bradlow, 2008; Hirata, 2004). We believe that these approaches are likely to be complementary.

To conclude, there are several suggestions for future studies based on the present first study on individual differences in the late acquisition of complex foreign consonant clusters. First, when studying the link between L1 skills and L2 learning abilities, a good, comprehensive measurement tool for L1 competence and a good, comprehensive way of assessing progress in L2 phonology are needed. Second, when generating test instruments to predict success in acquiring the phonology of a specific language, one would ideally show that performance in the training study is related to, for example, proficiency after 3 years of study. And third, given the large individual variation which was observed across all tasks, we conjecture that most likely only large-scale studies will provide reliable clues about the factors underlying performance differences in complex language learning tasks.

## Notes

- 1 We use the term L2 here in a general manner to refer to any second or foreign language acquired later in life after the first language has been established.

- 2 The control tasks were performed at the end rather than the beginning of the first session to avoid a possible influence on the participant's performance in the experimental tasks.
- 3 For further details on the Slovak phoneme inventory and the Slovak IPA transcription, see Hanulíková & Hamann (2010).
- 4  $A'$  and the associated measure of bias  $B''_d$ , are defined as functions of the number of "hits" (correct identification of a mispronunciation, for example) and "false alarms" (the incorrect reporting of a mispronunciation when none occurred), with  $A'$  reflecting the signal versus noise in the participant's responses and  $B''_d$  quantifying the participant's bias in answering (if any). For further details on how to calculate  $A'$  and  $B''_d$ , see Pallier (2002).
- 5 The only significant learning effects in paired t-tests were in the phoneme monitoring task between sessions 2 and 3:  $t(37) = -2.61, p = 0.013$ , and in the word monitoring task between sessions 1 and 2:  $t(39) = -2.37, p = 0.023$ .
- 6 Note that the results of the last session of phoneme monitoring are missing for subjects 6 and 39 due to technical problems.
- 7 The original data is plotted as box plots in Figure Supp. 1, see <http://sites.google.com/site/adrihanulik/home/supplement>.
- 8 The results showing these rank percentages for the participants in the top 20% (better than 80% of their peers) for at least 5 tasks can be seen in Table Supp. 1 at <http://sites.google.com/site/adrihanulik/home/supplement>.

## References

- Abu-rabia, S., & Kehat, S. (2004). The critical period for second language pronunciation: Is there such a thing? *Educational Psychology, 24*, 77–97.
- Altenberg, E. P. (2005). The judgment, perception, and production of consonant clusters in a second language. *IRAL (International Review of Applied Linguistics in Language Teaching), 43*, 53–80.
- Bent, T. (2005). *Perception and production of non-native prosodic categories*. Doctoral dissertation. Northwestern University.
- Birdsong, D. (2004). Second language acquisition and ultimate attainment. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 82–105). London: Blackwell.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*, 9–49.
- Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 99–116). Amsterdam: Benjamins.
- Boersma, P. (2001). Praat: A system for doing phonetics by computer. *Glott International, 5*, 341–345.

- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.
- Booij, G. (1995). *The phonology of Dutch*. Oxford, UK: Oxford University Press.
- Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). Philadelphia: John Benjamins.
- Bradlow, A. R., Akahane-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 & Psychophysics*, *61*, 977–985.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.
- Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics*, *2*, 5–14.
- Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–113). Rowley, MA: Newbury House.
- Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test (MLAT)*. San Antonio, TX: The Psychological Corporation.
- Darcy, I., Park, H., & Yang, C. (2011a). Individual differences in acquisition of L2 phonology. Talk presented at the Second Language Studies Colloquium, Indiana University, February 11th.
- Darcy, I., Park, H., & Yang, C. (2011b). Suppression of L1 influence on L2 phonological processing: Cognitive abilities and individual variation. Talk presented at Second Language Forum, Ames, Iowa, October 14th.
- Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*, 272–288.
- Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass*, *5*, 126–139.
- de Jong, K. J., Hao, Y. C., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, *37*, 357–373.
- de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, *91*, 450–476.
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453–482.

- Diaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *PNAS*, *105*, 16083–1688.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*, 35–39.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.
- Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). London: Blackwell.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589–1608.
- Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–272). Timonium, MD: York Press.
- Flege, J. E. (2005). Origin and development of the Speech Learning Model. Talk presented at the 1<sup>st</sup> ASA Workshop on L2 Speech Learning, Simon-Fraser University, Vancouver, BC. April 14–15.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.
- Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1–34
- Flege, J. E., MacKay, I., & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.
- Flege, J. E., Yeni-Komshian, G., & Liu, H. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104.
- Fogarty, J., Baker, R. S., & Hudson, S. E. (2005). Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. *Human Computer Interaction: Proceedings of Graphics Interface (GI 2005)*, 129–136.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Golestani, N., & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*, *17*, 929–934.
- Guiora, A. (1990). A psychological theory of second language pronunciation. *Toegepaste Taalwetenschap in Artikelen*, *37*, 15–23.

- Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, 24, 592–608.
- Hanulíková, A., & Dietrich, R. (2008). Die variable Coda in der slowakisch-deutschen Interimsprache. In M. Tarvas (Ed.), *Tradition und Geschichte im literarischen und sprachwissenschaftlichen Kontext* (pp. 119–130). Bern, Switzerland: Peter Lang.
- Hanulíková, A., & Hamann, S. (2010). Illustrations of Slovak IPA. *Journal of the International Phonetic Association*, 40, 373–378.
- Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrast in words versus sentence contexts. *Journal of the Acoustical Society of America*, 116, 2384–2394.
- Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-/l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, 15, 255–274.
- Kabak, B., & Idsardi, W. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, 50, 23–52.
- Leather, J., & James, A. (1996). Second language speech. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 269–316). San Diego, CA: Academic Press.
- Lee, S., & Cho, M. (2005). Perception and production of consonant clusters by Korean learners of English. *English Language and Literature*, 50, 1101–1132.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 6, 707–710.
- Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. *Proceedings of the 13th International Congress of Phonetic Sciences. Vol. 4.* (pp. 92–99). Stockholm, Sweden.
- Logan, J. S., Lively, E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of Acoustical Society of America*, 89, 874–886.
- Magnuson, J.S., Yamada, R. A., Tohkura, Y., Pisoni, D. B., Lively, S. E., & Bradlow, A. R. (1995). The role of talker variability in non-native phoneme training. In *Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan* (pp. 393–394). Tokyo: Acoustical Society of Japan.
- McLaughlin, B. (1990). The relationship between first and second languages: Language proficiency and language aptitude. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 158–178). Cambridge, UK: Cambridge University Press.
- Pallier, C. (2002). Computing discriminability and bias with the R software. Available at <http://www.pallier.org/ressources/aprime/aprime.pdf>.
- Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, 48, 98–107.

- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215.
- Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46–73.
- Rochet, B. L. (1995). Perception and production of L2 speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 379–410). Timonium, MD: York Press.
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3, 243–261.
- Shibuya, Y. (2005). Production and perception of consonant clusters in the L2 phonology of Japanese learners of English. Ph.D. dissertation. Georgetown University.
- Shin, D., & Iverson, P. (2011). Individual differences in vowel epenthesis among Korean learners of English. *Proceedings of the International Congress of Phonetic Sciences*, 1822–1825.
- Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, 51, 381–398.
- Sparks, R., & Ganschow, L. (1993). The effects of a multisensory structured language approach on the native language and foreign language aptitude skills of at-risk learners: A follow-up and replication study. *Annals of Dyslexia*, 43, 194–216.
- Sparks, R., Ganschow, L., & Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology*, 87, 638–655.
- Sparks, R., Ganschow, L., Patton, J., Artzer, M., Siebenhar, D., & Plageman, M. (1997). Prediction of foreign language proficiency. *Journal of Educational Psychology*, 89, 549–561.
- Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press.
- Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, 36, 131–145.
- Strange, W., Polka, L., & Aguilar, M. C. (1989). Effects of auditory and phonetic training on Americans' discrimination of Hindi retroflex-dental contrasts. *Journal of the Acoustical Society of America*, 86(Suppl.), S101.
- Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language*, 77, 647–723.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America*, 113, 1033–1043.

## Appendix A

### Experimental Items

Items session 1 & 2		Items session 3	
<i>words</i>	<i>nonwords</i>	<i>words</i>	<i>nonwords</i>
brnkút'	prnkút'	člnky	člngy
čln	člň	hrněk	chnřek
hrběk	chrběk	krstná	krstná
krb	grb	mrmloš	mrml'oš
krčma	krčna	pltník	bltník
krst	klst	prst	brst
míkvý	mígvý	pstruh	pstluh
mrkva	mrkfa	sklznica	sklsnica
plnka	plňka	škrtnút'	šklnút'
plst'	plšt'	slnko	srnko
plstnatý	plstmatý	smrdiet'	smldiet'
poštrngat'	poštngat'	smrt'	šmrt'
škrt	škl	spln	splň
škvrna	škvlna	srst'	srst
smrkat'	snrkat'	stlpec	štlpec
šmrnc	šmrně	stlstnutý	stlstmutý
šplh	splh	štrbské	strbské
stlmit'	štlmit'	strhnút'	strechnút'
stlpcový	stlpcový	štvt'	stvt'
strč	strc	svrbl'avé	svrblavé
štrk	strk	tvrdza	dvrđza
stlřpnút'	štřpnút'	vrzgnút'	frzgnút'
štvtřka	stvtřka	vtlkat'	vtřgat'
tlstý	trstý	vzhl'ad	vzhlad
tvrdnút'	dvrđnút'	zatrpknút'	zadrpknút'
vlnky	flnky	zhrbený	zhrpený
vtlčet'	vtlčet'	žltnút'	žrtnút'
žgrloška	žgrl'oška	zmlknút'	zmlkmút'
zmrřit'	smrřit'	zmrznút'	zmrzňút'
zvlhnút'	zvlhmút'	zvlhčit'	svlhčit'

## Appendix B

### Dutch and Slovak Phonology

SLOVAK	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p b		t d		c	k g	
Nasal	m		n		ɲ		
Fricative		f v	s z	ʃ ʒ		x	ɦ
Affricate			ts dz	tʃ dʒ			
Trill			r				
			r:				
Approximant					j		
Lateral approximant			l		ɫ		
			l:				

The table above shows the IPA chart for Slovak consonants (for more details, see Hanulíková & Hamann, 2010). There are only four Slovak phonemes (/r: l: ɫ /) that do not occur in Dutch as either phonemes or allophones or in loanwords. In Dutch, the non-native phonemes /g ʃ ʒ/ occur in loanwords and/or as allophones (e.g., Booij, 1995), and thus some familiarity with these phones can be assumed. Similarly, the Dutch alveolar phonemes /s z n/ are palatalized before /j/ and realized as postalveolar or prepalatal sounds /ʃ ʒ c ɲ/. Although these are contextually driven allophones only, they are not unfamiliar to Dutch speakers.

There are some differences in the orthography-phonology mapping. For example, Dutch <g> is pronounced /x/, but in Slovak as /g/ (or /k/ if devoiced). Furthermore, Dutch speakers have to learn that different diacritics indicate either length or palatalization in Slovak.

Some phonological processes are similar in both languages. Both Dutch and Slovak have regressive assimilation, thus the first consonant in a cluster of obstruents assimilates in voicing to the following consonants. Word finally obstruents are always voiceless in both languages.

Both languages have relatively complex syllabic structures (for more details, see Booij, 1995; Hanulíková & Dietrich, 2008). While Slovak phonology allows the occurrence of consonant clusters of up to four consonants in onset position, Dutch onsets consist of up to three consonants (e.g., Booij, 1995). Dutch syllables exhibit complex codas with up to five consonants, while more than two consonants in the Slovak syllable coda are rare and occur only in loanwords.

Importantly, there are several language-specific restrictions on the possible combination of segments (for more details, see Booij, 1995; Hanulíková & Dietrich, 2008), which might pose some challenges for the Dutch participants. In Slovak, short and long laterals and short and long rhotics can form syllabic peaks, which create seemingly vowelless words such as *prst* “finger.”

## Appendix C

### Questionnaire (translated from Dutch)

1. Gender: male female
2. Birthplace and year: \_\_\_\_\_ – \_\_\_\_\_ – \_\_\_\_\_  
City Country Year
3. Do you have hearing problems? yes no
4. What is your mother tongue? (multiple answers possible)  
Dutch  
Other [\_\_\_\_\_]
5. What languages were spoken at home when you were little (e.g., by your parents, grandparents or a babysitter)?  
\_\_\_\_\_
6. What languages do you understand without difficulty and speak them fluently?  
\_\_\_\_\_
7. What languages did you learn at school (or at university)?  
a. \_\_\_\_\_ How many years? \_\_\_\_\_ At what age? \_\_\_\_\_  
b. \_\_\_\_\_ How many years? \_\_\_\_\_ At what age? \_\_\_\_\_  
c. \_\_\_\_\_ How many years? \_\_\_\_\_ At what age? \_\_\_\_\_
8. In which country have you lived longer than 6 months?  
Country \_\_\_\_\_ from \_\_\_\_\_ to \_\_\_\_\_  
Country \_\_\_\_\_ from \_\_\_\_\_ to \_\_\_\_\_  
Country \_\_\_\_\_ from \_\_\_\_\_ to \_\_\_\_\_
9. Do you play a musical instrument or sing in a choir? If so, please specify.  
Yes, often \_\_\_\_\_  
Yes, sometimes \_\_\_\_\_  
No (go to question 12)
10. If you play an instrument or sing, please indicate at what age you started and how many years of uninterrupted lesson you have had.

Instrument: \_\_\_\_\_ Age: \_\_\_\_\_ Years of practice:  
\_\_\_\_\_

Instrument: \_\_\_\_\_ Age: \_\_\_\_\_ Years of practice:  
\_\_\_\_\_

Singing (level): \_\_\_\_\_ Age: \_\_\_\_\_ Years of practice:  
\_\_\_\_\_

11. How often do you practice?
  - Never
  - 0–1 hour per week
  - 2–6 hours per week
  - 7–10 hours per week
  - more than 10 hours per week
12. How often do you do sports per week?
  - Never
  - 1–2 times a week
  - 3–4 times a week
  - More than 4 times a week, I am a professional in \_\_\_\_\_
13. How good are you in sports in general?
  - excellent
  - good
  - not so good
14. Do you craft / draw?
  - Never
  - Often, I'm good in \_\_\_\_\_
  - Often, I have excellent skills in \_\_\_\_\_
15. How familiar are you with the language that you heard in the experiment?
  - Not at all. I have never heard it.
  - A little bit. I hear it occasionally.
  - Reasonable familiar. I hear it sometimes.
  - Familiar. I hear it regularly.
  - Very familiar. I hear it often.
16. Could you guess which language was spoken in the experiment?
  - Yes \_\_\_\_\_
  - No
17. Have you ever had lessons in phonetics / phonology?
  - Yes
  - No
18. What are you studying / did you study? \_\_\_\_\_