The role of explicit knowledge and experience with accent in the acquisition of second language sounds

Nikola Anna Eger

München, 2019

The role of explicit knowledge and experience with accent in the acquisition of second language sounds

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
der Ludwig-Maximilians-Universität München

vorgelegt von
Nikola Anna Eger
aus München

München, den 12.10.2018

Referentin: PD Dr. habil. Eva Reinisch

Koreferent: Prof. Dr. Jonathan Harrington

Datum der mündlichen Prüfung: 13.2.2019

DANKSAGUNG

Zuallererst möchte ich meiner Betreuerin Eva Reinisch meinen Dank aussprechen, für die wertvollen Ratschläge, ihre Hilfestellung und ihre unermüdliche Unterstützung.

Dann möchte ich mich bei Jonathan Harrington für die hilfreichen Fragen und Kommentare zu den Experimenten bedanken, die ein äußerst wichtiges Feedback waren.

Ein ganz besonderer Dank gilt meinem „Doktorbruder“ Miguel Llompart, für all die Gespräche und den guten Austausch in den letzten drei Jahren. Miguel, ich wünsche Dir von Herzen alles Gute für die Zukunft.

Holger Mitterer möchte ich ganz herzlich für seine Expertise und wertvollen Ideen bei den letzten Experimenten danken.

Darüber hinaus möchte ich einigen Personen für ihren Rat und ihre Unterstützung in meiner Zeit am IPS ganz besonders danken, darunter Phil Hoole, Wolfram Ziegler und Felicitas Kleber. Bei dem gesamten Institut für Phonetik und Sprachverarbeitung möchte ich mich für all das bedanken, was ich dort gelernt habe.

Bei Thomas Kisler, Markus Jochim, Katrin Wolfswinkler, und Lia Saki Bucar Shigemori möchte ich mich für den guten und wichtigen Austausch in den letzten Jahren bedanken. Ganz besonders danke ich Thomas Kisler für die Motivation auf den letzten Metern.

Bei Rosa Franzke möchte ich mich für ihre Mithilfe beim Rekrutieren, Testen, Auswerten und für den Austausch von Ideen und Probanden bedanken, und für ihre herzliche Art.

Zuletzt möchte ich mich bei meiner Familie und meinen Freunden bedanken, für ihre unerschöpfliche Geduld und ihren Beistand, ganz besonders bei Lorenzo und bei meinen Eltern für ihre Nachsicht und die Kraft, die sie mir geben.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Learning foreign languages (L2) has become an indispensable part of education these days. To exemplify, in 2015 more than 98 percent of lower secondary school pupils between about 11 and 18 years in the European Union studied at least one, and almost 60 percent studied two or more foreign languages (Eurostat, European Union, 2017). However, even when having started lessons early at school and even after many years of practice, L2 speakers often retain at least a detectable accent when they speak. Interestingly, sometimes they notice typical accent properties in fellow learners, for instance a German learner of English saying "[sɛ]nk you", but their own productions are accented in a very similar way. In other words, even though L2 learners notice the accent in other speakers of the same first language (L1), they seemingly do not use this knowledge to improve their own pronunciation. Whereas a wide range of studies has addressed the factors affecting the strength of a foreign accent, relatively little has been done to understand why it is so difficult to overcome this accent over time. The present thesis investigates some factors that may contribute to foreign accents being so persistent.

When henceforth the term *foreign accent* is used, this will refer to "patterns of speech resulting from L1 influence on the L2 that are noticeably different from native-speaker productions" (Derwing & Munro, 2015: 177). These patterns can be found in various dimensions of speech, on the segmental as well as on the suprasegmental level. On the latter, the word stress may not be on the correct syllable, or the speech rhythm and global intonation structure of a sentence may differ from how a native speaker would produce it (Bissiri & Pfitzinger, 2009; de Mareüil & Vieru-Dimulescu, 2006; Huan & Jun, 2011). On the segmental level, segments may be deleted, added, or substituted (e.g., Kubozono, 2002; Magen, 1998; Smith, Hayes-Harb, Bruss, & Harker, 2009). One example of this would be the German speaker mentioned above, who replaced the dental fricative /θ/ in the word *thank* with /s/, and produced the vowel /æ/ as a somewhat higher vowel /ɛ/. Foreign accent is typically characterized by a combination of all these aspects. Among all the various characteristics of foreign accent, this thesis focuses on the segmental level of accented speech. Thereby, both production and perception in a second language will be investigated.

There are two particularly famous models on L2 sound acquisition that are both based on the phonetic properties of the segments that occur in a language: The Speech Learning Model (SLM: Flege, 1995; 2003) and the Perceptual Assimilation Model (PAM-L2: Best

& Tyler, 2007). This thesis does not aim at testing these models, but they offer a good basis to describe the mechanism behind L2 production and perception. Both models acknowledge that, in general, there is evidence that the later in life learners start to learn a foreign language, the stronger their accent will be (e.g., Asher & García, 1969; Flege, 1995; Flege, Munro, & MacKay, 1995; Munro & Mann, 2005; Piske, MacKay, & Flege, 2001). The ability to acquire a good pronunciation in an L2 decreases gradually with increasing age at which an individual begins to learn that language, rather than changing abruptly (Flege, 1995; cf. Lenneberg, 1967; Patkowski, 1990). Other psycho-social factors have also been shown to affect the strength of an accent, such as the length of residence in a community in which the L2 is spoken, amount of L1 and L2 use, type of instruction, or motivation (Derwing, Munro, Foote, Waugh, & Fleming, 2014; Gluszek, Newheiser, & Dovidio, 2011; Ingvalson, Holt, & McClelland, 2012; Moyer, 2007). While both models admit that these factors can affect the overall accent, they argue that within a group of learners of the same L1, not all L2 sounds are equally easy to acquire.

The reason for this can be found in L1 acquisition: During the first year of life, infants perceptually tune into the sound system relevant for their own first language (Maurer & Werker, 2014; Werker & Tees, 1984). That is, they establish distinct mental representations of the sounds they are exposed to, which become more and more defined in acoustic space and also in terms of motor goals (Guenther, 2006; Tourville & Guenther, 2011). In the acquisition of a foreign language, when the L1 categories are already in place, all new sounds are subsequently perceived through a native filter or "sieve" (Trubetzkoy, 1939: 47). Both SLM and PAM-L2 take this L1-filtered perception into account: They predict that how well a specific L2 sound will be acquired depends on its articulatory and perceptual (dis-)similarity to native sound categories.

Especially at the beginning of L2 acquisition, unfamiliar sounds that are similar but not identical to L1 categories are likely to be perceived as the closest native sound that already exists in the L1. Depending on the language and sound combinations, there may happen to be two L2 sounds that both map onto one single L1 category. For these sounds, the models predict that they will not be well distinguished and perceived as very similar. For instance, the English front vowels /æ/ and /ɛ/ are produced with slightly different tongue heights and therefore differ in acoustic/spectral properties and perceived sound quality (e.g. Ladefoged & Maddieson, 1996). Since German has only one vowel in the front mid-open space, German learners should be poor at perceiving the difference between the two vowels. Because the vowel that exists in German is similar to /ɛ/, the unfamiliar vowel

/æ/ should be particularly difficult. Indeed, research on various L1-L2 combinations has demonstrated the difficulty that learners have in perceptually differentiating vowel or consonant contrasts of this kind (e.g., Best, McRoberts, & Goodell, 2001; Cebrian, 2000; Cutler, Weber, & Otake, 2006; Hattori & Iverson, 2008; Llompart & Reinisch, 2017; Sheldon & Strange, 1982). Moreover, recent research indicated that difficult L2 sounds, like English /æ/ for German learners, are not represented as equal to the closest L1 sound in the mental lexicon. Rather, they are stored as an unspecific or fuzzy version of it (Cutler, 2015; Cutler et al. 2006; Darcy, Daidone & Kojima, 2014; Escudero, Hayes-Harb, & Mitterer, 2008). That is, learners can develop mental representations of L2 sounds, but these are initially shaped by their L1-filtered perception, and thus overlap at least partially with L1 categories. Those sounds that are difficult to perceive are typically the ones that also challenge learners in production, as will be taken up below in some more detail (e.g., Bohn & Flege, 1992).

L2 sound contrasts can also be present in the L1, but may be unfamiliar in a specific word position. English has a phonemic voicing distinction for stops and fricatives, which can occur in word-initial, word-medial (intervocalic), and word-final position. Word-finally, such as in *pick-pig* or *leaf-leave*, the contrast can be signaled by the voicing of the consonant itself, but also by the temporal relation between the consonant and the preceding vowel: A relatively shorter vowel combined with a longer consonant signals a voiceless stop or fricative as in *pick* or *leaf*, whereas a relatively longer vowel and shorter consonant signal a phonemically voiced consonant as in *pig* or *leave* (Port & Dalby, 1982; Wright, 2004). In German, the voicing distinction in stops and fricatives exists, but word-finally, all obstruents are canonically devoiced[1]. When German learners produce English words ending in voiced obstruents, they often transfer this canonical devoicing from their L1 to the L2, thereby minimizing the L2 contrast (Smith et al. 2009). In both the voicing contrasts in word-final obstruents and the vowel contrast discussed above, one category is familiar because it occurs in L1 (/ɛ/ and the voiceless obstruents) while the other is unfamiliar as category or in a specific position (/æ/ and the phonemic voiced obstruents). This kind of sound contrasts will be of interest in Chapter 2 and 3.

Another case can be made with sounds or sound pairs that are difficult because they are acoustically and articulatorily different from any native sound category. A case in point

[1] There is evidence that this neutralization of the final devoicing is incomplete in German, but the contrast is not robust in production and not used in perception in a straightforward way (Roettger, Winter, Grawunder, Kirby, & Grice, 2014; see also Kleber, John, & Harrington, 2010).

is the glottal fricative /h/ in English or German for learners with, for instance, French or Italian as L1. Since there is no obvious L1 sound that could be used instead, one might assume that in this case learners just ignore the unfamiliar sound in perception, and delete it in production[2]. Contrary to this, a study on French learners of German showed that learners did produce the /h/ in about 70 %. When /h/ was not produced appropriately, the preferred strategy was not to completely delete it, but to produce a glottal stop or glottalization instead (Zimmerer & Trouvain, 2015). Glottal stops or glottalization can be used as a stress marker in French (Malécot, 1975). Earlier studies on the perception of sounds that are very dissimilar from any native category have shown that learners can be surprisingly good at discriminating them (e.g., Best, McRoberts, & Sithole, 1988). However, in order to fully master an L2 contrast, learners do not only have to perceptually discern the sounds, they also need to encode them correctly in the lexical items they belong to. That is, they additionally need to assign the sounds correctly to words (e.g., Hayes-Harb & Masuda, 2008). Especially with regard to sounds that are dissimilar form any native category, little is known on how these are used for spoken-word recognition in more natural listening situations. Therefore, this issue will be addressed in Chapter 4.

Difficulties with a specific L2 contrast are typically found in both perception and production, indicating that there is a link between these two modalities. Assuming this, there are different possibilities how this relation may look and how the two modalities may develop in an L2. One possibility is that a good pronunciation of difficult L2 sounds has to be preceded by a good perception of those (Flege, 1995). Alternatively, good production skills may enhance or strengthen perceptual skills in a second language. Prior research on this link revealed mixed results: Some studies provide evidence that production benefits from better perception (e.g., Sakai & Moormann, 2018; Underbakke, 1993), whereas others indicate that perception can be preceded by production (Sheldon & Strange, 1982; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005). Moreover, some studies found correlations between production and perception abilities in an L2 (group-wise or at an individual level), while others found only weak or no correlations at all (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Wong, 2013). In sum, although there is evidence that production and perception abilities in an L2 are not entirely separate, the link between them is not yet fully understood.

[2] Note that the SLM predicts that the greater the perceived phonetic dissimilarity between an L2 sound from a native category, the more likely a new category will be established (Flege, 2003).

Producing one L2 sound inappropriately – for example pronouncing the word *pig* as sounding very similar to *pick* – typically comes at the cost of maintaining a relevant L2 contrast. That is, speaking with a foreign accent can diminish the acoustic differences between relevant L2 contrasts, but that does not always mean that the contrast is not maintained at all (e.g., Hanulíková & Weber, 2012). Moreover, there is usually more than one way to signal a contrast. L2 contrasts are often maintained in a nonnative manner, frequently because learners transfer cues from the L1 that are only secondary or irrelevant to native listeners (e.g., Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; but see also Bohn, 1995; Cebrian, 2000). For instance, to differentiate the unfamiliar vowel contrast /ɛ-æ/, German learners may use vowel duration, an acoustic cue they are familiar with in their L1, whereas native English speakers rely primarily on spectral cues, that is, vowel quality (Bohn, 1995). In addition, there are also individual differences between learners regarding how well they can maintain the contrast and in which cues they use to maintain it (Schertz, Cho, Lotto, & Warner, 2015; 2016; Wade, Jongman, & Sereno, 2007). In sum, learners often minimize L2 contrasts in production and perception, mainly because they use cues to a smaller extent than native speakers and in an untypical manner. Foreign accent thus reflects nonnative-like perception and production of L2 sounds and contrasts that result from differences in the phonetic-phonological systems between L1 and L2.

Now, what does this mean for interactions between native and nonnative speakers? Considering the phonetic-phonological characteristics described above, native listeners should have more problems[3] in understanding foreign-accented than native speech: Phonemes are used to differentiate meaning between words within the phonological system of a language. Replacing one sound by another can change the meaning of a word or render it a non-word. Such sound substitutions can frequently be found in foreign-accented speech, as described above: Producing the vowel /æ/ as sounding very similar to /ɛ/ turns *pan* to *pen*, and *dragon* to dr[ɛ]gon, a form that does not correspond to a specific lexical item in native English. Indeed, a study on American-accented Dutch revealed that Dutch listeners were worse at identifying words spoken by American learners than when produced by other

[3] Note that accent can affect both listeners' intelligibility as well as comprehensibility, which is the *experienced* ease or difficulty in understanding an intended message. However, even strongly accented speech can be fully intelligible, but it may be more demanding to do so. Hence, there is a relation between accent, intelligibility and comprehensibility, but this relation is complex and the domains are partly independent (Derwing & Munro, 2015).

native speakers of Dutch. Crucially, this was primarily due to the learners' pronunciation of Dutch vowels that do not occur in their L1 (van Wijngaarden, 2001).

The initial difficulties that native listeners have when trying to understand foreign-accented speech can be overcome after a phase of exposure. That is, after being presented with even short samples of accented speech, listeners can "tune into" a specific accent and become better at understanding it. This can be explained by listeners' flexibility in adapting their sound representations according to the current input: With regard to segments, listeners temporarily adjust the phoneme boundaries between two sounds when exposed to non-canonical pronunciations (e.g., Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009; Witteman, Weber, & McQueen, 2013). For instance, when native listeners are presented with words in which the vowel /æ/ is pronounced as [ɛ], they may broaden their perceptual category of /æ/ and consequently perceive or accept more vowels as belonging to that category, even if their acoustic properties are more /ɛ/-like. This effect is known as perceptual recalibration (for an overview, see Samuel & Kraljic, 2009).

However, not all listeners have the same initial problems in understanding foreign-accented speech. Nonnative listeners of the same L1 as the speaker can be equally good at understanding accented and non-accented speech (Bent & Bradlow, 2003) or be even better when listening to accented speech (Xie & Fowler, 2013). To exemplify, Spanish learners of English were better able to understand a text read by another Spanish learner compared to when it was read by a native speaker of English (Major, Fitzmaurice, Bunta, & Balasubramanian, 2002). Moreover, nonnative listeners can be better than native listeners at understanding foreign-accented speech (Hayes-Harb, Smith, Bent, & Bradlow, 2008; Xie & Fowler, 2013). Even though this effect does not exist in an all-or-nothing fashion (e.g., Stibbart & Lee, 2006), it has repeatedly been shown that nonnative speakers have fewer problems to understand accented speech than native listeners. This phenomenon has been termed interlanguage intelligibility benefit, and it has been explained by "shared phonetic and phonological knowledge" (Bent & Bradlow, 2003: 1607) between speaker and listener. However, there is evidence that additionally, the benefit could be a result of long-term exposure to an accent (Li & Mok, 2015; Xie & Fowler, 2013). This is a reasonable explanation for at least two reasons: First, listeners tune into non-canonical pronunciation variants independently of their L1, that is, even if they do not share phonetic-phonological knowledge with the speaker (Weber, DiBetta, & McQueen, 2014). Secondly, adaptation due to frequent exposure seems plausible given that foreign languages are often acquired,

at least initially, in a classroom situation. There, learners are frequently confronted with accented speech from their fellow learners or teachers.

Therefore, if exposure to foreign-accented speech shapes the mental representations, accented speech should be not only more intelligible, but also sound more *acceptable* for learners compared to native listeners. This issue will be addressed in Chapter 2 by retrieving explicit judgments on the pronunciation of words from minimal pairs with difficult contrasts for German learners of English: The vowel contrast between /ɛ/ and /æ/ and the voicing distinction in word-final fricatives and stops. The combination of German as an L1 and English as an L2 offers a good opportunity to test the question whether foreign-accented productions may be perceived as acceptable instances of L2 words. In German schools, teachers are rarely native speakers of the target language, and during their education, they are often not obliged to spend time abroad to practice the language they are going to teach. It is hence not unusual that they have limited exposure to native English and speak with a German accent. Moreover, movies and series in German TV are dubbed, and only in the past few years it has become more common to watch films in the original language. That is, in contrast to other countries like the Netherlands, the Scandinavian countries, or Finland, Germans are less likely exposed to native speech[4], including English (Koolstra, Peeters, & Spinhof, 2002). Importantly, while German learners of English are probably all exposed to German-accented English, contact with English spoken by native speakers naturally varies among learners. Therefore, the experiment will also address the role of the listeners' experience and usage of English.

German learners of English and English native speakers have to rate the goodness of English words that were produced by another group of German learners. The presented words differed in how accented they were according to whether they contained a familiar, "easy", or an unfamiliar, "difficult" sound (which is expected to be produced less accurately and sound more accented), and according to the overall production quality. Importantly, and contrary to previous studies, it is not accent ratings that are retrieved (Flege, Munro, & MacKay, 1995; Thompson, 1991). Instead, listeners have to rate the *goodness* of the pronunciation. This type of task is chosen in order to tap into the learners' lexicon: When judging how well a word was pronounced, listeners have to compare the version heard to some stored representation of it. If an accented word is rated as "good", this suggests a

[4] Note that whether the media has an influence on pronunciation in L1 and L2 is still an ongoing debate (e.g., Rindal, 2010; Stuart-Smith, 2007; Trudgill, 2014).

perceived match with a stored representation. The use of word-sized material is hence important to investigate how *words* may be represented in the mental lexicon.

The first hypothesis is that if familiarity with the accent of one's own L1 shapes the mental representations of L2 words, German learners should be less sensitive to German accent than native speakers of English. Moreover, if contact to native speech has an impact on mental representations, German learners with more exposure to English should be more sensitive to German accent (that is, the differences between words with easy and difficult sounds and according to overall production quality), and in that more similar to the native listener control group. Results can hence inform us about whether familiarity with the accent of one's own L1 affects how L2 words are stored in the lexicon, and how this is modulated by contact to native input. Judgements on the goodness of foreign-accented words are likely related to having an accent: If an accented utterance is perceived as a good instance of an English word as learners have become used to it, there is no obvious need to change pronunciation.

Chapter 2 will hence investigate the influence of familiarity with accent typical of the learner's own L1 on the perception of L2 words. If it is the case that familiarity with specific production patterns plays a role for the mental representation of words, there is another important aspect to be considered: the voice that one listens to (our own voice vs. voices of others). The role of one's own voice is a unique one for two main reasons. First, it may be the case that the "person to whom we listen most is ourself" (Levelt, Roelofs, & Meyer, 1999: 6). Secondly, speakers are at the same time always listeners of themselves. There is hence a special connection between what is planned (forward models), what is actually being said (speech movements), and what is received (auditory and sensory feedback; Tourville & Guenther, 2011). Therefore, Chapter 3 will address the role of one's own voice in a second language.

One's own voice is perceived differently from others' voices, due to different conduction ways over ear plus bone for one's own vs. ear only for other voices. Nevertheless, speakers are able to recognize themselves when presented with a recording of their own voice (e.g., Rosa, Lassonde, Pinard, Keenan, & Belin, 2008; see also Shuster & Durrant, 2003). Previous research has shown that the perception of one's own voice differs from the perception of others as indicated by various behavioral and (neuro)physiological measures (e.g., Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008). For instance, listeners are better at recognizing their own voice than another, familiar one, especially under difficult listening conditions (Xu, Homae, Hashimoto, & Hagiwara, 2013).

A study that examined how brain activity changes as response to different voices has found that automatic attention was reduced when one's own voice was perceived compared to other voices (Graux, Gomot, Roux, Bonnet-Brilhault, Camus, & Bruneau, 2013). This has been interpreted as a mechanism for communication, where it is particularly relevant to pay attention to others. Importantly, self-other differences have also been found with respect to phonetic-phonological awareness: Children with a phonological disorder were worse at correctly identifying their own erroneous productions than those of other children (Shuster, 1998; see also Strömbergsson, Wengelin, & House, 2014). One explanation for this may be that children's representations of the sounds they have difficulties with are imprecise because they are repeatedly exposed to their own erroneous speech. This finding is especially interesting because phonological difficulties and segmental deviations are also found in foreign-accented speech.

Interestingly, there seem to be both enhancement and suppression mechanisms when perceiving one's own voice: On the one hand, the *recognition* of one's own voice is enhanced compared to other people's voices, on the other, *attention* to one's own speech and *awareness* of one's errors seem to be reduced. This may indicate that better familiarity with one's own voice and lowered attention to one's own productions are possibly two sides of the same coin. What could this mean for the acquisition of a second language? As argued above, L2 sound contrasts are diminished by learners, in production and perception, but they are often not completely neutralized. Moreover, learners differ considerably in how well they produce L2 sounds and contrasts, but also according to *how* they realize them. Given this variability, the question arises as to whether learners – due to experience with own accented production patterns – are better at recognizing L2 words with difficult L2 contrasts that were spoken by themselves than productions of others, even if the contrasts are similarly well produced.

This question will be addressed in two experiments reported in Chapter 3. In the first experiment, German learners of English perform a word recognition task on the same difficult English minimal pairs as in Chapter 2. Crucially, participants are also presented with their own productions of these words among the productions of several other unfamiliar speakers who produced the key contrasts in a similar way. The first hypothesis is that if familiarity to own production patterns in an L2 influences word recognition, learners should be better at understanding L2 words produced by themselves than words produced by other, unfamiliar learners. A word recognition task was chosen, instead of a goodness rating task as in Chapter 2, in order to avoid the influence of positive or negative

attitudes towards own pronunciation in an L2 when one's own voice is recognized, as self-conception in general differs considerably among individuals (e.g., John & Robins, 1994). A word recognition task avoids the influence of the listeners' attitude towards one's own voice on goodness ratings, but it can show whether listeners are better at perceiving their own productions compared to other speakers' utterances. Moreover, previous studies on L2 word-identification have demonstrated that the above-mentioned intelligibility benefit is observed particularly for low-proficiency listeners when presented with productions of low-proficiency speakers (Hayes-Harb et al. 2008; Xie & Fowler, 2013). Building on these findings, a second hypothesis was that lower-proficiency learners may benefit most from hearing their own voice. If a self-benefit can be found, this indicates that familiarity with one's own speech patterns affects mental representations and word recognition in an L2. A better recognition of self-produced words, however, may go hand in hand with a reduced awareness of one's own accent: Own productions may be better recognized *despite* having been produced inappropriately, that is, with an accent.

The second experiment in Chapter 3 focuses on the production-perception link in L2. Findings on whether and how L2 production and perception abilities are linked are diverse. A weak point of many studies is that the methods to compare production and perception use very different types of stimuli, for instance accent ratings of learners' productions on one hand, and learners' discrimination of synthetic stimuli on the other (e.g., Flege, Bohn, & Jang 1997; Hattori & Iverson, 2010). The second experiment in Chapter 3 will address this issue by using the same stimuli to examine learners' production and perception abilities. The questions are whether L2 learners with better production skills outperform learners with poor production skills also in perception, and whether this depends on the quality of the material. Results from these experiments can give insights in whether mental representations of L2 words are shaped by one's own production patterns and how this is related to L2 proficiency. A self-benefit in recognition of L2 words may indicate that learners adapt over time to their own accented speech patterns. Adaptation and good recognition despite a foreign accent, in turn, may be detrimental to improve pronunciation, as learners may not experience a need to change.

If the results from Chapter 2 and 3 indicate that representations of L2 sounds and words are shaped by foreign accent, the reason is likely to be frequent exposure to accented speech patterns, and limited experience with native speech. This process typically takes place over a longer period of time, and may be both automatic and unconscious. However, there are some aspects in a second language learners are aware of, for instance,

orthography. Chapter 4 will address the role of explicit knowledge of a difficult L2 sound, as mediated by orthographic coding, on how well it is acquired. To investigate this, the combination of Italian as an L1 and German as an L2 offers a good opportunity: The glottal fricative /h/ is a well-known example of a difficult L2 sound for learners whose L1 is French or Italian, for instance, as the phonological systems of these languages lack this sound category. In German, the glottal fricative contrasts with the glottal stop, which is canonically produced in the beginning of words that are orthographically vowel-initial, such as *Apfel* "apple". Even though the two sounds share some properties, for instance a restricted distribution (Wiese, 1996), and both have been found to be of importance for native speakers of German in spoken word recognition (Mitterer & Reinisch, 2015), they differ substantially in their status. /h/ is coded orthographically and learners typically know it as a difficult sound. For /ʔ/, there is no letter in German, and learners (as well as native speakers) are usually not aware that this sound exists. The experiments reported in Chapter 4 investigate how Italian learners of German produce and perceive these two sounds, which do not occur as speech segments in Italian except for in hyper-articulated speech, paralinguistic function or to signal stress or phrase boundaries (Bertinetto & Loporcaro, 2005; Stevens, Hajek, & Absalom, 2002). If explicit knowledge of a new sound has an impact on its acquisition, /h/ should be better acquired than /ʔ/.

There is only little research on the acquisition of L2 sounds that have no clear counterpart in the L1, and among the existing studies only few focus on lexical processing. In general, listeners have been shown to perform better in phonetic tasks that focus on acoustic differences, whereas they may perform poorly in tasks that tap into lexical processing. In other words, even though learners may be able to acoustically differentiate two L2 sounds, they do not necessarily allocate them correctly in words to be used in more natural speaking and listening situations (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Llompart & Reinisch, in press_a, in press_b; Sebastián-Gallés & Díaz, 2012). Therefore, the experiments reported in Chapter 4 investigate production, as well as perception in an implicit and a more explicit task. For the explicit task, learners' goodness judgments of words starting with /h/ or /ʔ/ in their correct form are compared to either words with sound substitutions (e.g., [ʔ]ut for *Hut* "hat", and [h]apfel for *Apfel* "apple"), or deletions. For the implicit perception task, the same material is used but this time in an eye-tracking experiment. By using this method, eye-movements to written words or depicted objects can be measured in real time. It offers an excellent possibility to test spoken-word recognition in more natural settings, where listeners spontaneously react to

speech input by looking at visual referents. Assuming that listeners' reactions reveal how well an auditorily presented word matches a potential lexical candidate, this method gives insights into how mental representations may be shaped and how they are used to access lexical entries (Huettig, Rommers, & Meyer, 2011). Findings from these experiments can hence contribute to the existing literature in that they inform us on how two L2 sounds that have no clear counterpart in the L1 are produced and perceived. Their contribution is also to test whether explicit knowledge helps establish a new category, and how this may be beneficial in explicit tasks and spoken word processing.

In sum, the main objective of this thesis is to investigate why it is so difficult to overcome one's accent in a foreign language. Chapters 2 and 3 address the role of familiarity with an accent: Chapter 2 asks whether familiarity with the accent typical of one's own L1 shapes representations in a way that accented words sound acceptable to learners. Chapter 3 investigates whether familiarity with one's own very specific production patterns in an L2 leads to better word recognition of own than others' productions. Chapter 4 focuses on the influence of explicit knowledge of an L2 sound's existence, and how this may affect production, perception in an explicit task and spoken-word processing. Results from these experiments can hence give insights into whether familiarity with accented production patterns and awareness of L2 sounds shape the learners' mental representations of L2 sounds and words.

2 ACOUSTIC CUES AND PROFICIENCY IN ACCENT PERCEPTION

Abstract

The speech of second language learners is often influenced by phonetic patterns of their first language. This can make them difficult to understand, but sometimes for listeners of the same first language to a lesser extent than for native listeners. The present study investigates listeners' awareness of the accent by asking whether accented speech is not only more intelligible but also more *acceptable* to nonnative than native listeners. English native speakers and German learners rated the goodness of words spoken by other German learners. Production quality was determined by measuring acoustic differences between minimal pairs with "easy" vs. "difficult" sounds. Higher proficient learners were more sensitive to differences in production quality and between easy and difficult sounds, patterning with native listeners. Lower proficient learners did not perceive such differences. Perceiving accented productions as good instances of L2 words may hinder development since the need for improvement may not be obvious.

A version of this chapter was published in the Journal of *Studies in Second Language Acquisition* (Eger & Reinisch, 2019b)

2.1 INTRODUCTION

Learners of a second language (L2) have to overcome many challenges, among many others, to accurately perceive and produce words that contain difficult L2 sounds. For example, German learners of English struggle to differentiate the vowels in word pairs such as *pen* vs. *pan* (Llompart & Reinisch, 2017). As a consequence, they are often perceived to speak with a foreign accent. Foreign-accented speech usually deviates from how native speakers of the target language would typically speak, and is therefore often more difficult to understand than native productions, for native and nonnative listeners (Imai, Flege, & Walley, 2003; van Wijngaarden, 2001). However, to L2 learners, foreign-accented speech can sometimes be as intelligible as native, non-accented speech (Bent & Bradlow, 2003), specifically when listener and speaker share the first language (L1) background. This benefit has been proposed to arise from shared knowledge about the phonetics and phonology of the learners' L1. Additionally, it could result from long-term exposure and hence adaptation to accented productions. This is likely, considering that many L2 learners learn their second language in a classroom situation where they have ample experience with nonnative speech from their classmates and often also from the teacher. If learners were exposed to and adapted to accented speech from the onset of learning, for them accented speech may not only be as intelligible but also as *acceptable* as native speech of the target language because the accented forms may have become a good fit to the representation of these words. As a consequence, learners may be less aware of the accent of their L1 than native listeners of the target language. In the present study, we asked whether German learners of English indeed perceive English words spoken with a German accent as more acceptable instances of these words the lower their own proficiency and experience with English. Results will be compared to native speakers of English.

Native listeners are usually quite good at detecting a foreign accent in another talker's speech, even when presented with short utterances or single words (Flege, 1984). This is because nonnative productions differ along many dimensions from native speech, for example, the word stress may not be on the correct syllable, the temporal relation between sounds may differ from a native manner, sounds are substituted with others, or differ in sub-segmental detail (e.g., Bent, Bradlow, & Smith, 2008; Bissiri & Pfitzinger, 2009; Smith, Hayes-Harb, Bruss, & Harker, 2009; Wester, Gilbers, & Lowie, 2007). Foreign accent is usually characterized by a combination of all these aspects. It has been shown that developmental and socio-psychological factors are important determiners of the strength of a learner's accent, for instance, age of learning, length of residence in the L2

environment, the amount of first and second language use, or motivation, to name but a few factors (for recent overviews see, e.g., Gluszek, Newheiser, & Dovidio, 2011; Ingvalson, Holt, & McClelland, 2012; Moyer, 2007; Piske, MacKay, & Flege, 2001).

However, from a linguistic point of view, whether or not a given L2 sound will be easy or difficult to learn also depends on the phonetic and phonological properties of the learner's first language sound inventory compared to the L2 that should be learned (Best & Tyler, 2007; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008). Models of second language acquisition (e.g., PAM-L2: Best & Tyler, 2007; SLM: Flege, 2003; NLM-e, Kuhl et al. 2008) propose that the ease with which a separate representation for a new L2 sound can be established, depends on how distinct the new sound is to the closest L1 categories. A new L2 sound contrast is especially difficult to learn (in both, perception and production) when the two L2 categories are perceptually mapped onto a single native category. Then learners also tend to produce the L2 contrast less distinctively and less consistently than native speakers (e.g., Levy & Law II, 2010; Smith et al. 2009; Wade, Jongman, & Sereno, 2007). That is, even if a learner can distinguish between the sounds of a new L2 contrast, the cues they use in perception and production may differ from native speakers of the target language (Escudero, Benders, & Lipski, 2009; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Levy & Law II, 2010; Schertz, Cho, Lotto, & Warner, 2015). Since in addition, L2 speech is often characterized by large inter- and intra-speaker variability (Wade et al. 2007), native listeners tend to show more difficulties in understanding and slower processing of foreign-accented speech than non-accented speech (Ferguson, Jongman, Sereno, & Keum, 2010; Munro & Derwing, 1999; van Wijngaarden, 2001).

Despite initial difficulties in understanding accented speech, it has been shown that listeners are able to quickly adapt to non-canonical productions such as found in foreign-accented speech (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Reinisch & Weber, 2012; Sidaras, Alexander, & Nygaard, 2009; Witteman, Weber, & McQueen, 2013). That is, already after brief exposure to accented speech listeners become better and faster at recognizing words or sentences spoken with a previously unfamiliar accent. Importantly, adaptation does not only occur in an experimental setting, but also through “natural” experience with accented speech outside the laboratory (Witteman et al. 2013). In a priming study, Witteman et al. (2013) showed that Dutch listeners who had every-day experience with German-accented Dutch were better able to process German-accented words than listeners with limited experience that they accumulated over the course of the experiment

(see also Sebastián-Gallés, Echeverría, & Bosch, 2005). Moreover, Dutch listeners who were familiar with an Italian accent showed facilitation in understanding Italian-accented Dutch as well as Italian-accented English words. That is, adaptation occurred in, or transferred to a second language (Weber, Di Betta, & McQueen, 2014; see also Reinisch, Weber, & Mitterer, 2013).

Critically, when listening to accents in a second language listeners are often better able to recognize words if the accent in the stimuli matches the accent of their own L1 (Bent & Bradlow, 2003; Weber, Broersma, & Aoyagi, 2011; Xie & Fowler, 2013). For example, Bent and Bradlow (2003) showed that for Korean learners of English, Korean-accented English was as intelligible as native, non-accented English, even if the Korean accent was defined as strong. That is, the learners had a benefit insofar as that they did *not* have more difficulties in understanding English spoken in their own accent compared to native, non-accented English. This was in contrast to native English listeners who clearly understood accented speech less well than native speech. Moreover, in a similar type of study, Spanish speakers of English were better able to answer questions after listening to a lecture that had been read by Spanish speakers of English compared to when read by native English speakers (Major, Fitzmaurice, Bunta, & Balasubramanian, 2002). However, in the same experiment, the other tested language groups, Japanese and Chinese learners of English, did not show such an advantage for their own L1-accent. They were similarly good or better when listening to native speakers of English (Major et al. 2002; see also Munro, Derwing, & Morton, 2006). Hayes-Harb, Smith, Bent and Bradlow (2008) suggest that the interlanguage intelligibility benefit holds specifically for poor learners and when listening to poorly pronounced words. Harding (2012) adds that the benefit may be task-dependent. However, although the interlanguage intelligibility benefit may not be an all-or-nothing phenomenon, tendencies for an advantage for understanding one's own familiar L1 accent have repeatedly been found. This issue will be taken up in the discussion of the present results.

Importantly, when looking for a possible explanation of such a benefit, when observed, it has been suggested that it comes from knowledge about the phonetics of the learners' first language. Since L1 phonetic and phonological patterns often affect the pronunciation of L2 speech sounds, listeners whose L1 corresponds to the accent in the speech sample may have an advantage over other listeners. If in addition, learners have ample experience with the accent of their L1 it could be assumed that for them overall familiarity with the accent may also add to their ease of understanding.

Adaptation to accented speech and subsequent benefits in the speed and accuracy of recognizing accented words have been demonstrated for native and nonnative listeners (e.g., Reinisch & Weber, 2012; Sebastián-Gallés et al. 2005; Sidaras et al. 2009; Witteman et al. 2013). The more the listeners had adapted, the more accurately they recognized words and the more quickly this happened. What remains unanswered is whether foreign-accented productions also sound better to the listener when asked explicitly. This is especially likely if an L2 is learned in an L1 environment where learners have ample exposure to accented speech. If as a result of adaptation, accented productions were not only well intelligible but also *acceptable* forms of the target words, this could suggest that accent has become part of the learners' representations of the L2 (see e.g., Cutler, 2015, for a discussion of L2 lexical representations). That is, accented forms may have become a reasonably good match to listeners' reference representations because listeners are familiar with common forms of mispronunciations as possible pronunciation variants of the target words. For example, German learners of English who often produce English words like *birthday* as "bir[s]day" with an "s" instead of "th" and frequently hear this form produced by fellow learners, may accept bir[s]day as a possible or even reasonably good form of *birthday* (Hanulíková & Weber, 2012).

Critically, if learners judge accented words as acceptable instances of the target form, this may have consequences for their own improvement in the L2 since the need for a change may not be obvious. Note that there is some prior evidence that listeners who are familiar with an accent are less harsh in judging this accent (Schmid & Hopp, 2014; Thompson, 1991; Winke, Gass, & Myford, 2013). It has been proposed that listeners' judgments of a foreign accent become harsher, once they become sensitive to phonetic divergences from non-accented forms. Only with longer experience, the perceived strength of the accent reduces again, suggesting adaptation (Flege & Fletcher, 1992).

In the present study, we asked how German learners of English at different levels of proficiency and with different amounts of exposure would rate the quality of German-accented productions. We presented native English listeners and German learners of English with German-accented words that varied in the magnitude of deviation from typical English productions. We asked German learners as well as native English listeners how well they thought these words were produced. In contrast to other studies that investigated the perceived strength of the accent (e.g., Munro et al. 2006), we specifically asked listeners to rate the *goodness* of a produced word. In this way, we aimed to tap into the learner's explicit knowledge of target form: When judging how well a word is pronounced, the

listener has to compare the word to some inner representation of it. If a word was rated as well pronounced, this would suggest that there was a perceived "match" with a stored representation of this word in the learner's mental lexicon. To minimize possible influences of suprasegmental aspects of the accent we focused on single, monosyllabic words containing sounds from difficult sound contrasts. As mentioned above, the pronunciation of certain nonnative sounds is one relevant factor that contributes to a perceived foreign accent and at least native listeners have been shown to detect foreign accent reliably even in short utterances (Flege, 1984).

Specifically, we investigate two types of English sound contrasts that have been shown to be difficult for German learners. The vowel contrast /ɛ/ – /æ/ (see, e.g., Bohn & Flege, 1992; Llompart & Reinisch, 2017) and the word-final voicing contrast in obstruents (Smith et al. 2009). As for the vowel contrast, German, unlike English, has only one lax open mid-front vowel[5], which is acoustically and articulatorily close to English /ɛ/. Therefore, this vowel is usually easy for Germans to perceive and produce. The other somewhat more open English mid-front vowel category /æ/ does not exist in German. German learners often have difficulties to perceptually discern it from /ɛ/ and as a consequence often also produce it as /ɛ/-like. This pronunciation may be mistaken as the other vowel by native English listeners, that is, an intended production of *pan* may be perceived as *pen*. /æ/ is hence a difficult sound for Germans. A similar case can be made for the word-final obstruents. In German, there is a phonemic contrast between /b,d,g,z,v/ and /p,t,k,s,f/ in word-initial and -medial position, but unlike in English it is neutralized word-finally[6]. German learners of English often transfer this neutralization in favor of the voiceless sounds to English (Smith et al. 2009). Thus, words ending in a voiced stop or fricative, like *pig*, are more "difficult" for Germans, whereas words like *pick* are rather "easy".

The main aim of the present study was to test how German learners of English perceive German-accented words depending on their own English proficiency. Since we expected that the more proficient learners are in their L2, the closer their behavior would

[5] Orthographically there is also a tense vowel <ä> but its phonemic status in contemporary spoken German is unclear. In many German varieties, it is pronounced as /e:/ and a pronunciation with a more open vowel is marked as very clear speaking style (Becker, 2012) or as part of certain dialects (e.g., Alemannic dialects, see Hobel, Moosmüller, & Kaseß, 2016).

[6] Note that this neutralization in German has been shown to be incomplete (e.g., Roettger, Winter, Grawunder, Kirby, & Grice, 2014). However, German listeners do not use this information directly in perception (Kleber, John, & Harrington, 2010).

be to that of native listeners, we also included a native-listener reference group. The perception of accent was tested by asking how well learners would perceive differences in production quality of accented words, and specifically between words with easy and difficult sounds since the latter are more likely to be produced with an accent.

Our first expectation was that learners with higher proficiency in English will be more likely to perceive a difference in goodness of pronunciation between words with easy and difficult sounds compared to lower-proficient learners. In other words, learners with lower proficiency and less practice in English should be less sensitive to an accent in fellow learners' productions. As concerns the quality of the tokens, we expected that the better the tokens were produced, the better they would be rated overall. Moreover, the perceived difference in goodness between easy and difficult sounds would be larger in overall poorly produced tokens. This is because in poor productions the difficult sound may be perceived as clearly worse than the easy sound. Again, we asked to what degree listener proficiency would modulate this effect. By specifically testing the relations between the factors Sound Type (easy vs. difficult sounds), production quality of the tokens ("Material": good, intermediate and poor productions) and listener Proficiency (learners of different levels of proficiency and a native listener reference group), the present study set out to test learners' perceptual sensitivity to accent in L2 productions. Focusing on accent that matches the listeners' L1, we would like to speculate that perceiving accented productions as good instances of L2 words may affect initial L2 development since the need for improvement may not be obvious.

2.2 METHOD

2.2.1 Participants

Twenty monolingual native speakers of English and thirty German learners of English participated for pay. They reported no history of speech, language or hearing problems. The native English speakers were undergraduate college students at the University of California, Berkeley (henceforth "American listeners") aged between 18 and 23. None of them spoke German or had contact with German learners of English. The German learners of English were students at the University of Munich, Germany. Their mean age was 25.2 years (sd = 3.1) ranging from 20 to 33. All speakers had learned English at school in Germany starting at an average age of 10.0 years (sd = 1.9, with the youngest starting at 5 and the oldest at 13 years) where they followed classes for an average of 8.7 years (sd =

1.6, ranging from 6 to 12 years). Participants were selected such that they would be representative of typical German learners of English who had not spent more than 6 months in an English-speaking country. Four of the 30 participants reported to have spent some time in a country which is dominantly English-speaking but for less than half a year. At the time of the experiment, all German participants lived in Germany and used English only according to personal habits ranging from hardly any use at all to moderate contact through the internet (note that films and series on German TV are dubbed into German). This information was assessed in a questionnaire asking about habits of usage of English and self-rated proficiency.

In order to test whether the German learners' proficiency in English as a second language influences how they perceive German-accented English, a score was calculated based on five dimensions from the questionnaire. Note that our use of the term "proficiency" does not refer to the number of years of learning English but rather to a combination of usage-based factors: Specifically, the first two dimensions refer to self-reported frequency of speaking and listening in English. Additionally, the learners' self-estimated speaking skills and self-estimated proficiency in listening comprehension in English were considered. As a fifth dimension the learner's self-estimated accent when speaking English was included. Each question could be answered on a seven-point scale, with 1 indicating frequent use, good skills or weak accent, and 7 indicating infrequent use, poor skills or strong accent, respectively. The mean of the five responses was calculated so that each participant received one value that represented his or her "proficiency".

2.2.2 Materials

Thirty-one English minimal word pairs were selected that differed in sound contrasts that have been shown to cause problems for German learners in production and perception (Llompart & Reinisch, 2017; Smith et al. 2009). Eleven minimal pairs were chosen to differ in the vowel contrast /ɛ/ – /æ/, seven pairs in the word-final voicing contrast in fricatives and thirteen pairs in the word-final voicing contrast in stops. Within each pair, one word contained sounds that had been shown to be "easy" for German learners. These were the /ɛ/ in words such as *pen*, and the word-final voiceless stops or fricatives in words such as *pick* or *rice*. The other word of the minimal pair contained a sound that had been shown to be "difficult" for German learners. These were the vowel /æ/ like in *pan* and word final voiced stops or fricatives, such as in the words *pig* or *rise*. As described in the introduction, the labels "easy" and "difficult" were based on whether or not the critical sounds occur in

the German sound inventory (German does not have the vowel /æ/) and in the given word position (German word-final phonologically voiced obstruents are canonically produced as devoiced). Words containing either an easy or a difficult sound will be henceforth termed easy or difficult word, respectively. An additional 22 words were selected to serve as fillers for the recording session. Words are listed in Appendix I.A.

For the recordings, all words were randomly assigned to one of ten semantically neutral carrier sentences such as *The next word is...* . Target words were always in sentence final position. The order of words was randomized with the restriction that the words of a minimal pair could not follow one another. Each word was repeated twice for a total of 160 sentences[7].

Twenty-four female[8] German learners of English were recorded of which later a subset was selected to represent a range of different proficiency levels. Speakers were recruited according to the same criteria as reported in the section "participants" above, but none participated later in the main accent-rating experiment. Speakers were instructed in English and asked to read out the entire sentence at a comfortable pace. The sentences including the target word were presented one by one on a screen. The recordings were made in a soundproof recording room using a diaphragm microphone (Neumann Microphone, type TLM 103) and *Speechrecorder* software (Draxler & Jänsch, 2004), which stored each sentence as a separate wav file on a computer.

A subset of speakers was selected to form a representative sample of different proficiency levels, four speakers per group A, B, or C (A=best, B=intermediate, C=worst). The assignment was done separately for each sound contrast and based on how well a given speaker had produced a given critical sound contrast. To assess this production "quality" and to select speakers, acoustic analyses were conducted on the productions of all speakers.

Several acoustic measures were taken for all 24 speakers for each sound contrast using Praat (Version 5.4.08, Boersma & Weenink, 2015). For the vowels, these were the first two formants and duration; for the word-final fricatives, these were the duration of the preceding vowel and the duration of the fricative (combined as vowel duration divided by fricative duration) and the voiced portion of the fricative; and for the word-final stops, the duration of the aspiration, the duration of the preceding vowel and the voiced portion of

[7] The words *bet*, *bat*, *bed* and *bad* were used for the stop voicing contrast as well as for the vowel contrast, but each word was recorded only twice.

[8] Only female speakers were recruited to focus listeners' attention on the pronunciation of the critical words/sounds rather than differences in voice (quality).

the closure. These acoustic measures were selected because they have been shown to be the most important cues to the respective contrast for native speakers and listeners of English (see e.g., Deterding, 1997; Hillenbrand, Getty, Clark, & Wheeler, 1995, for the vowels; e.g., Broersma, 2010; Wright, 2004, for the fricatives; e.g., Barry, 1979; Smith et al. 2009, for the stops). A good contrast was defined as a large difference between the means of the acoustic measures for the two categories across words. Cues to each contrast were weighted in the order named above. First, tokens of the eight speakers who had produced the clearest contrasts of the learners were assigned to group A. Then, the eight speakers with the smallest produced contrasts were assigned to group C. The remaining eight speakers were assigned to group B. Since this assignment was done separately for each sound contrast and in order to reduce the overall number of speakers for the perception experiment, a subset of four speakers per contrast per proficiency group was selected. Overall, productions from 13 different speakers were included (i.e., one speaker could be used for more than one sound contrast).

Note that in the remainder of the paper we will refer to the variable of speaker proficiency with the label Material in order to not confuse it with proficiency of the listeners in the perception task. Material has the levels A, B, and C, where A tokens had been produced most clearly (i.e., larger mean differences and more cues to differentiate the words of the minimal pair), and C tokens showed only a small mean difference and more overlap between the words of the minimal pairs. Tokens from set B were intermediate. The main acoustic measures for each type of contrast and the three material sets can be seen in Figures I.B.1 to I.B.3 in the Appendix.

2.2.3 Design

For the goodness rating task, the words of the minimal pairs spoken by the selected speakers were spliced out of the carrier sentences to be presented in isolation. To further reduce the number of trials presented in the experiment, one of the recorded repetitions per word and only five word pairs per contrast type were selected (see Appendix I.A). The selection proceeded as follows: first, words with other difficult sounds than the critical contrast were excluded (e.g., words with the contrast /ɛ/-/æ/ that happened to end in a voiced obstruent). Second, words were excluded for which more than two of the speakers indicated that they did not know the meaning (as assessed in a questionnaire after the recordings). The final set of stimuli consisted of 2 words x 5 pairs x 3 sound contrasts x 4 speakers per contrast x

3 speaker groups (Material sets A, B, and C) for a total of 360 trials and was the same for all listeners.

2.2.4 Procedure

The English listener group participated at the University of California, Berkeley, in the United States. The German listener groups participated at the University of Munich in Germany. All participants received written instructions in English. For the Germans, this was to set them into an English language mode without influencing their perception by talking to them with a specific accent. The written instructions, the material and the procedure were the same for both listener groups.

Participants were seated in a sound-proof booth in front of a laptop computer. On each trial, they saw one word of the minimal pair in orthographic form in the middle of the computer screen and below a five-point scale with the labels "very good" and "very poor" at the end points. After 300 ms the target word was presented over headphones at a comfortable listening level. The participants' task was to indicate how well the word was pronounced by pressing one of the number keys from 1 to 5 on a standard computer keyboard. Five hundred ms after the response was recorded, the next trial started automatically. All words in the perception task formed minimal pairs with another word according to one of the three critical sound contrasts. However, at any given trial throughout the experiment only one word was presented at a time auditorily and orthographically. The written word always matched the intended form of the spoken word (i.e., it matched the word that speakers had read during the recordings). For half of the participants in each group the response key *1* was labelled "very good" and *5* "very poor", whereas for the other half the labels were reversed. The numbers of the scale were always ordered from left (1) to right (5). The words were presented in randomized order, and every 60 trials participants were allowed to take a self-paced break. The experiment was implemented in PsychoPy2 (Version 1.83.01; Peirce, 2007), and took approximately 15 minutes to complete.

2.2.5 Analysis

All statistical analyses were conducted in R (Version 3.3.2, R Core Team, 2017) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) using linear-mixed effects regression models. Mixed models have been shown to be preferable over traditional analyses of variance (ANOVA) in designs such as ours that have repeated measures over

participants and items. They are less susceptible to Type I errors in such cases (Quené & van den Bergh, 2008). Random effects take into account that participants and items may differ idiosyncratically and by estimating participant and item idiosyncrasies, they also allow an estimate how likely it is that the same result would be obtained if the experiment was repeated with different participants and items. Random effects subsume random intercepts and random slopes. Random intercepts estimate to what extent a given participant or item provided ratings above or below average, while random slopes capture differences in the sensitivity to fixed-factor effects (e.g., to what extent pronunciation ratings for an item are strongly or weakly influenced by the acoustic realization of the contrast; see e.g., Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Field, Miles, & Field, 2012, for more detailed discussions of mixed-effects models).

For the present analyses, two such linear mixed-effects models were run, one for analyzing the responses of the American listeners and one for the German learners. The dependent variable was the rating for a given word from a given speaker, recoded so that "1" always indicates that listeners rated the pronunciation of the presented word as "very poor" and "5" as "very good" with 2, 3, and 4 as intermediate steps. This rating was used as the dependent variable in both models.

For the model of the native listeners we analyzed two variables of interest and their interaction: Sound Type which referred to the "easy" (coded as 0.5) vs. "difficult" (coded as -0.5) sound within a given sound contrast, and Material. The latter referred to how well the contrast had been produced according to the acoustic measures discussed above (see also Appendix I.B). Material had three levels A, B, and, C (A = largest contrast/best production, B = intermediate, C = smallest contrast/worst production) that were coded as numeric with A=0.5, B=0, and C=-0.5. For the analysis of the German learners' responses, listener Proficiency was added as a third variable of interest along with all interactions with the other factors. Proficiency was calculated for each participant as the mean of five self-ratings from the questionnaire (on a scale from 1-7; see Participant section above). For the statistical analysis and Figure 2.1, these values were centered on the group mean and recoded so that they conform to a "higher-is-better" model of evaluations. With this coding, the grand mean is mapped onto the intercept, and effects and interactions can be interpreted similar to traditional ANOVA.

The random-effects structures for both models included random intercepts for participant and word (i.e., item) with random slopes for all fixed factors and their interactions that were manipulated within participants and items respectively (Barr et al.

2013; i.e., within participant: Sound Type and Material, within item: Material, and Proficiency in the case of the learner model).

In order to illustrate the statistically significant effects and interactions for the native listeners and the learners, as well as a descriptive comparison between the two listener groups two types of plots are presented in Figure 2.1. The three panels from left to right show listeners' ratings for the three Material sets A, B, and C. While the scatterplots in the upper panels focus on effects and interactions involving listener proficiency, the barplots in the lower panels zoom in on the effect of Sound Type.

The y-axis in the upper panels (scatter plots) indicates the difference between the ratings for the easy and the difficult words. That is, the higher the value the better the easy words were rated compared to the difficult ones. A value of zero means that both were rated as equally good. Hence, an effect of Sound Type would be reflected in values that differ from 0. The x-axis in the upper panels indicates the proficiency of the learners with native listeners added at the very right. As for the analyses, the learners' proficiency values are centered with higher values indicating higher proficiency. Additionally, regression-coefficients were calculated for the German learners for each Material set in order to estimate the strength of the interactions between listener Proficiency and Sound Type. Note, however, that these were calculated using linear regression for each of the Material subsets and without adding random effects (i.e., using the lm() function in the package ‘stats’ in R; R Core Team, 2017). The coefficients are given in Figure 2.1.

The y-axis in the lower panels (barplots) shows the mean ratings for the easy and difficult words with the factor Sound Type indicated in light vs. dark colored bars. Here the effect of Sound Type across Material sets can be appreciated more directly than in the upper panels. However, for this illustration listener proficiency has been collapsed into poor learners, good learners and native listeners. The German learners were grouped by a mean split (i.e., what would amount to value zero in the top panels).

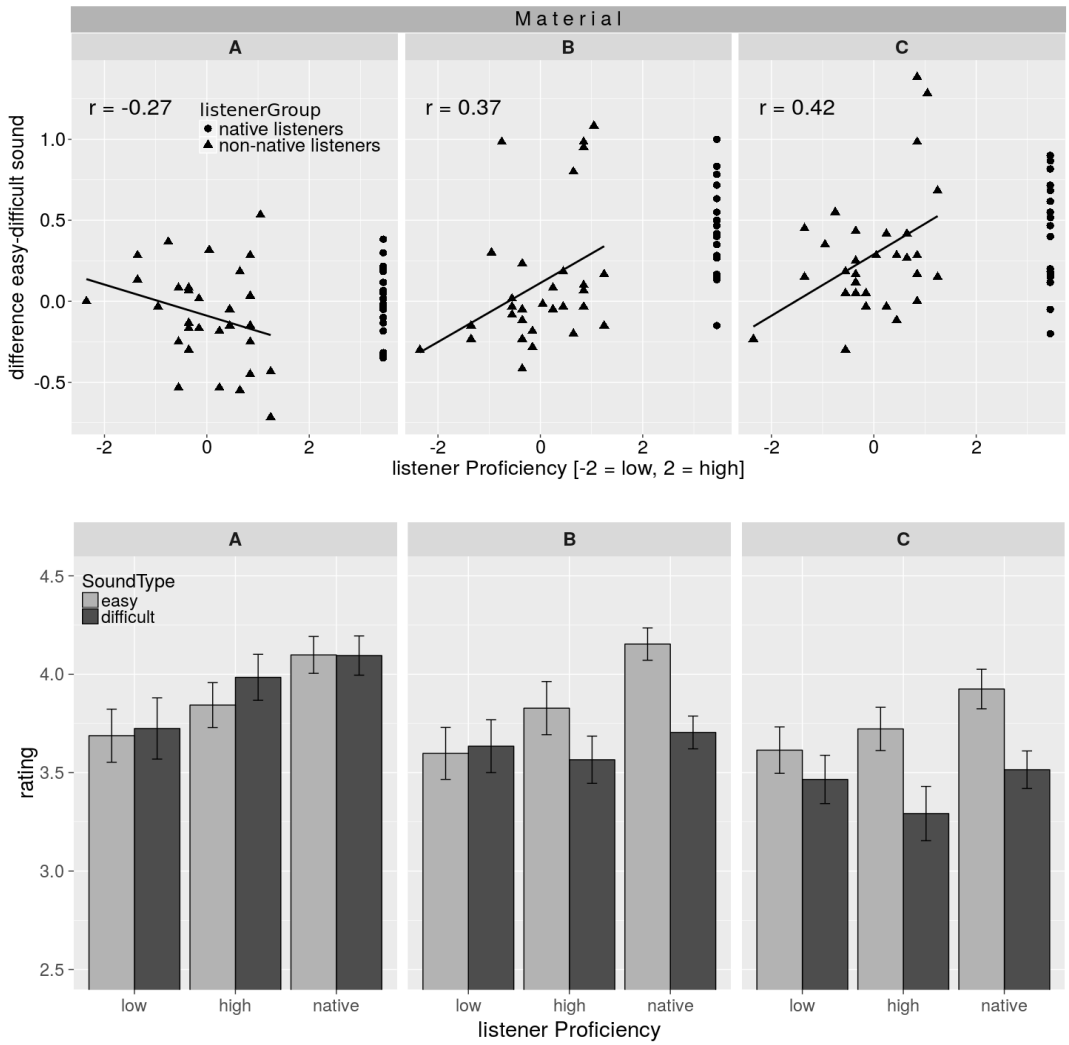

Figure 2.1: Means of listeners' ratings from 1 (very poor) to 5 (very good) presented in a scatterplot (upper panels) and in a barplot (lower panels). In the upper plot, the difference between ratings for easy and difficult sounds is shown for the three Material sets (A, B, C), for the range of listener proficiencies (-2 = low, +2 = high), and the native listeners at the very right. In the lower plot, the mean of listeners' ratings is shown for the three Material sets and for the two Sound Types (easy, difficult) separately. Here, listeners are grouped into low-proficiency German (left), high-proficiency German (mid), and American ("native", right). The German listeners are assigned to one of two proficiency groups by a mean split. Note that only the range from 2.5 to 4.5 of the responses is shown in order to better illustrate differences. Error bars represent 1 Standard Error and were adjusted for within-participant factors (see Morey, 2008).

2.3 RESULTS

2.3.1 Native listeners

A first overall model was fitted for the American listeners with the two factors Sound Type and Material, and interactions between them. This model served as a "baseline", to test our

basic assumption that easy words are rated as better than difficult words and that this may depend on the overall quality of the production.

Results show a significant effect of Sound Type suggesting that American listeners rated easy words better than difficult words (b=0.29, SE=0.12, df=35.25, t=2.45, p<.05; $b_{\text{Intercept}}$=3.92, SE=0.10, df=33.80, t=39.60, p<.001). Furthermore, there was an effect of Material (b=0.38, SE=0.08, df=30.54, t=5.00, p<.001) and a significant interaction between Sound Type and Material (b=-0.41, SE=0.16, df=33.05, t=-2.61, p<.05). Since the variable Material was coded as numeric with 0.5 for set A, the positive regression weight indicates that the better the tokens, the better ratings were given by the American listeners. The interaction indicates that the effect of Sound Type (better ratings for the easy than the difficult words) was larger the worse the Material set (in Material Sets B and C). This interaction is clearly visible in Figure 2.1. In the upper panels the difference between easy and difficult sounds in Material Set A is centered around zero (no difference between easy and difficult sounds) but clearly positive for Sets B and C (i.e., the easy sounds were rated better). The separate ratings for easy and difficult words and their interaction with Material are also illustrated by the bars in the lower panels. The results of this first model hence confirm that the assignment to material sets according to acoustically measured cues is reflected in the native listeners' ratings. As expected, native listeners perceived the accent stronger in the difficult than easy words. This effect becomes larger from the well to the poorly produced tokens, where the cues are less differentiated (from set A to set C).

2.3.2 German learners

The statistical model for the learners included the fixed factors Sound Type, Material, listener Proficiency and all interactions. Statistics are reported in Table 2.1. There was no significant effect of Sound Type, but a significant effect of Material indicating that the better the tokens, the better ratings the German listeners gave. However, Material was involved in several interactions. First, as for the native listeners there was an interaction between Sound Type and Material. Looking at Figure 2.1 this can be seen in that in the upper panels the difference between easy and difficult sounds is approximately centered around zero for Material set A (i.e., no difference) but moves towards positive values, that is, a larger difference, as the Material gets worse (i.e., towards C).

Importantly, the effect of Material as well as the interaction between Material and Sound Type was further modulated by listener Proficiency, as indicated in the two-way interaction between Material and Proficiency and the three-way interaction between all

three factors. The two-way interaction suggests that overall worse ratings were given from Material sets A to C the higher the listeners' proficiency. The three-way interaction suggests that the difference in ratings between easy and difficult sounds across material sets also depended on listeners' proficiency. This is illustrated in the scatterplots (upper panels of Figure 2.1) showing little change in the difference between easy and difficult sound as proficiency increases in Material set A (with a non-significant correlation in the opposite-than-expected direction). However, the difference in ratings for easy vs. difficult sounds increases the higher proficient the learners as we move to Material Sets B and C. This observation is confirmed by the regression-coefficients for interactions shown in the scatterplots, with a stronger correlation in Material C compared to B. The barplots in the lower panels of Figure 2.1 give a more direct impression of the effect of Sound Type across learners and Material sets. As can be clearly seen from both types of plots as well as the direction of statistically significant effects - the higher proficient the German learners the more they pattern with the native speakers.

Table 2.1: Results of the mixed-effects model fitted with Sound Type, Material, listener Proficiency, and their interactions for the German learners.

Fixed effect	*b*	SE	df	*t*	*p*
Intercept	3.66	0.08	35.33	43.85	<.001
Sound Type	0.10	0.08	49.70	1.31	=.20
Material	0.29	0.07	44.31	4.39	<.001
Proficiency	0.01	0.09	28.69	0.13	=.90
Sound Type:Material	-0.38	0.10	32.50	-3.66	<.001
Sound Type:Proficiency	0.09	0.07	32.51	1.35	=.19
Material:Proficiency	0.17	0.06	29.12	2.96	<.01
Sound Type:Material:Proficiency	-0.29	0.07	42.75	-4.08	<.001

2.4 DISCUSSION

The aim of the present study was to test how German learners of English judge the accent in English words spoken by other German learners, and whether they perceive accented productions as more acceptable instances of the intended English words than native English listeners do. This question was motivated by the observation that L2 learners often

understand foreign-accented speech just as well as non-accented speech, and in some cases, they also have an advantage over native listeners in understanding accented speech (e.g., Bent & Bradlow, 2003; Hayes-Harb et al. 2008; Imai et al. 2003). This benefit has been argued to result from shared phonetic and phonological knowledge about the speaker's first language. If, in addition, learners are frequently exposed to the L2 spoken with their L1 accent, accented productions may be picked up as possible variants to the intended words (Flege & Fletcher, 1992). If this was the case, words spoken with a foreign accent typical of the learners' own L1 should not only be as intelligible, but also as *acceptable* as non-accented productions. L2 learners may hence be less "sensitive" to differences in L2 productions than native speakers - specifically to differences between easy vs. difficult sounds, and, more generally, to differences in the quality of the productions. These hypotheses were tested with a group of German learners of English along a range of proficiencies who were asked to rate English words containing easy vs. difficult sounds spoken by other German learners of varying proficiency. The same type of ratings was obtained from a group of native English listeners from the US.

There were two main findings. First, the more proficient German learners of English are, the more sensitive they are to different degrees of accent in L2-productions of speakers of the same L1. This was the case for differences in easy vs. difficult sounds, as well as the overall quality of the tokens. Second, the higher the proficiency of the learners, the more similar their behavior is to the native listeners. Reversely, the less proficient learners are, the less sensitive they appeared to the strength of the accent in productions of learners with the same L1.

Note that our factor "Proficiency" was determined based on five dimensions from a questionnaire (see Method section) that focused on self-rated oral proficiency as well as self-reported frequency of use. The differences between learners could hence not be accounted for by factors such as length of learning or amount of instruction since all learners received instruction at school but not ever since then. Rather our proficiency variable was defined based on L2 use and included experience and practice of the L2 at the time of the experiment. Specifically, the more proficient learners also reported being regularly exposed to native English via television and the internet. The experience of learners with less frequent exposure was more likely to be limited to the lessons they had at school where their exposure was primarily to German-accented English.

In addition to testing listener proficiency, the present study set out to systematically test effects of the L2 material that listeners had to judge. Note that most previous studies

either focused on the learners' accents as rated by native listeners (Flege, Munro, & MacKay, 1995; Guion, Flege, & Loftin, 2000; but see Munro et al. 2006) or they focused on how well learners understand native English forms (Broersma, 2012; Weber & Cutler, 2004; but see Bent & Bradlow, 2003; Hayes-Harb et al. 2008; Weber et al. 2011). The material we used were words in isolation, specifically minimal pairs that differed in one critical sound contrast. In this way, the assignment of tokens to material sets could be based on acoustic measures. Importantly, results showed that differences according to these measures are reflected in the native listeners' ratings. Moreover, also learners showed sensitivity to the difference between easy and difficult sounds and to different degrees of accent (i.e., material), but this depended on their proficiency in the L2 (i.e., the three-way interaction). While higher-proficiency participants with more self-reported experience with native English patterned similar to the natives, participants with lower proficiency appeared to perceive little difference between the quality of productions.

We hypothesized that this could have at least two possible sources: the lower-proficiency listeners may not perceive the accent in the speakers' productions because the accent is based on an L1 phonology that corresponds to their own - as has been suggested for the interlanguage intelligibility benefit (Bent & Bradlow, 2003). Alternatively or additionally due to frequent exposure to the L1 accented forms, listeners became used to accented pronunciation and therefore accept the accented forms as reasonably good match to their reference representations.

Being asked what a speaker says, learners have repeatedly been shown to have less difficulties at understanding accented L2 speech compared to native listeners of the target language (e.g., Bent & Bradlow, 2003). However, in the present study learners had to explicitly rate how well a word was pronounced, which was known to the listeners as provided in its orthographic form. Whereas familiarity with a certain non-canonical pronunciation may be advantageous in a transcription or listening comprehension task, it may appear as a disadvantage when being asked to judge the strength of the accent. This may be because a "good" match could possibly be found even if the pronunciation differed from how a native speaker would produce the word - because learners have frequently heard accented variants. The finding that lower-proficiency learners appear to show no sensitivity to accent differences in other learners' productions, but higher-proficiency learners do, hence goes with the assumption that the less proficient learners are, the less native-like their representations of L2 words are. This finding is also in line with studies indicating that the interlanguage intelligibility benefit holds only for low-proficiency

learners (e.g., Hayes-Harb et al. 2008; Pinet, Iverson, & Huckvale, 2011; van Wijngaarden, Steeneken, & Houtgast, 2002; Xie & Fowler, 2013). For instance, Hayes-Harb et al. (2008) found a shared-L1 benefit for Mandarin learners of English only for low-proficiency listeners and if the material was produced by low-proficiency speakers. An acoustic analysis of the tokens that caused the largest benefit for low-proficiency listeners over native listeners revealed that the benefit has presumably been caused by a differential use of cues to the specific contrast (the word-final voicing contrast in stop consonants, which does not exist in Mandarin Chinese). Whereas native listeners were misled by the way the L2 speakers had produced the contrast, low-proficiency listeners of the same L1 interpreted the cues in the same nonnative way as the speakers, resulting in better recognition. The finding that this was true only for the low-proficiency listeners may indicate that the learners' representations are – at this stage of L2 acquisition – mainly shaped by their L1 accent. The more experience learners get with native cues to difficult L2 contrasts the closer their cue weighting may become to native speakers (though they may never fully match; Schertz et al. 2015; 2016). Also in the present study, the high-proficiency learners were sensitive to differences in acoustic characteristics of the accent, similarly to native listeners. The low-proficiency learners, by contrast, may have had advantage in word recognition due to a typically nonnative use of cues, and hence appeared "accent-deaf" when explicitly judging second language speech with accent that matches their L1.

More specifically, the lower-proficiency learners have likely established a representation of the target words that is somewhat "fuzzy" especially with regard to difficult sound contrasts (e.g., Darcy, Daidone, & Kojima, 2014; Weber & Cutler, 2004; see e.g., Cutler, 2015, for an overview). This fuzziness could be the result of difficulties in perceiving new L2 contrasts (Best & Tyler, 2007; Flege, 1995). Additionally, due to poor L1 accented input, representations are likely to be shaped in an even more nonnative way. Therefore, the mapping from the accented or native L2 signal is usually a good match.

Since the present study used an explicit goodness rating task with single words, the results could suggest that the inexperienced learners are less aware of an accent that corresponds to their first language than listeners with more practice in their L2. This interpretation is in line with previous studies using other types of material, for example, Munro et al. (2006) who showed that Japanese learners rated narratives in English produced by Japanese learners as less accented than English native listeners did. Reduced awareness may be one consequence of being mainly exposed to accented pronunciation variants. However, the awareness of accent may be one important factor in L2 pronunciation.

As concerns development in a second language, our results suggest that with more language experience and native input, representations of L2 words become more native-like. That is, even though learners may still be used to the accent of their L1 they are able to establish more target-like representations to which the accented input can be compared. Note that this development is expected and necessary since in many classrooms nonnative teachers have to grade students' productions. However, despite our finding that learners' behavior becomes more native-like with increasing L2 proficiency, the present results are not sufficient to tell *how* this transition from less to more experienced would proceed. Note also that L2 models assume that learners are able to change over time but leave the exact mechanisms for future research. A quantification of how much input is necessary for developing new or more target-like representations, however, is not trivial. A number of studies showed that additional information about differences between difficult L2 categories may help learners to start developing separate representations of these L2 sounds. This additional information can either be explicit instruction (such as corrective feedback, e.g., Saito & Lyster, 2012; Thomson, 2012; for an overview see Derwing & Munro, 2015, chapters 5 and 7) or when learning new words at a more advanced stage even implicit, for example, orthographic information or visible articulation (e.g., Escudero, Hayes-Harb, & Mitterer, 2008; Llompart & Reinisch, 2017). Future work will have to show how a combination of native-accented input, (meta)knowledge about L2 categories and awareness of a foreign accent influence how learners' abilities develop in a second language.

Conclusion

The present study showed that the more proficient and experienced L2 learners are in their second language the more sensitive they become to accent in L2 words produced by other learners of the same L1. They thereby appear to rely on similar acoustic cues as native listeners by specifically differentiating the production quality of easy vs. difficult sounds, that do not occur in their L1, and by differentiating different degrees of accent. Unlike that, listeners whose experience with spoken English is more limited to speech produced by speakers of the same L1, are more likely to accept accented productions as good instances of L2 words. We suggest that with more native input, representations can become less “accented” and more target-like. However, future research will have to show how learners can break out of the circle of perceiving the L2 through their L1 filter and compare new

input to accented representations. The ability to explicitly judge how well a word was pronounced may be one important aspect to start a change.

3 A SELF-BENEFIT FOR SPOKEN-WORD RECOGNITION IN L2

Abstract

Second language (L2) learners often speak with a strong accent, which can make them difficult to understand. However, familiarity with an accent enhances intelligibility. We propose that L2 learners are even more familiar with their own accented speech patterns and may thus understand self-produced L2 words better than others' accented productions, presumably due to adaptation. This hypothesis was tested by asking German learners of English to identify English words from minimal pairs that are distinguished by difficult L2 sound contrasts. Words had been spoken by the learners themselves or other learners who produced the contrasts equally well. Self-produced words were identified significantly better than others' productions. A second experiment revealed that better producers can exploit acoustic cues in perception more than poor producers, especially when the produced acoustic cues to the minimal pairs were clearly differentiated. The self-benefit, however, did not depend on production skills. We conclude that L2 learners adapt not only to their L1 accent in general but also to their own specific speech patterns. Speculating about L2 acquisition more generally, these results may raise the question whether adaptation to own, accented productions may be one reason why learners have difficulties to improve their pronunciation, since they may not notice a need to improve.

A version of this chapter was published in the *Journal of Experimental Psychology: Learning, Memory, and Cognition*. (Eger & Reinisch, 2019a).

3.1 INTRODUCTION

Learning a new language brings along many challenges. Even learners who have been using a second language (L2) for a very long time often retain a perceptible foreign accent in their L2 pronunciation. That is, their production of sounds and prosodic characteristics of the second language differs from how native speakers of that language would typically produce them. Interestingly, L2 learners are often well aware that their fellow learners (e.g., in the classroom) speak with a strong accent. Anecdotal evidence comes from the many jokes about learners' failure to produce certain sounds correctly (e.g., a German sailor saying "[s]ink positive"). However, given this awareness of others' errors, the question arises as to why listeners would not (always) be able to use that kind of information to improve their own accent.

One possible reason for this failure may be because L2 learners are highly familiar with their own accents. Familiarity with foreign-accented speech in general is beneficial to understanding the accent. The other side of the coin to also better understanding one's own accent, however, may be reduced awareness of one's own errors. In the present study, we test the part of this suggestion that L2 learners understand their own speech better than the speech of other L2 learners, presumably as a consequence of greater experience with and exposure to "self" speech compared to "other" speech. This seems plausible since the perception of others has been shown to differ in several aspects from the perception of oneself. Differences in the perception of self vs. others have been shown with regard to face recognition (see e.g., Devue & Brédart, 2011, for an overview), body (e.g., Ionta, Gassert, & Blanke, 2011), odor (e.g., Platek, Burch, & Gallup, 2001), and, importantly, voice and word recognition (e.g., Douglas & Gibbins, 1983; Schuerman, Meyer, & McQueen, 2015; Shuster, 1998; Xu, Homae, Hashimoto, & Hagiwara, 2013).

The sound of our voice that we are used to hearing while speaking differs from the sound we hear on a recording. This is due to the different routes via which the sound is conducted: air vs. air and bone (Shuster & Durrant, 2003). However, people have been shown to be good at recognizing their own voice on recordings and differentiate it from unfamiliar voices as indicated by a variety of behavioral and physiological measures (Aruffo & Shore, 2012; Douglas & Gibbins, 1983; Graux, Gomot, Roux, Bonnet-Brilhault, Camus, & Bruneau, 2013). Graux and colleagues, for example, used event-related potentials (ERPs) to show that neural processes involved in discriminating one's own voice from others' voices differs from processes involved in the discrimination of two unknown voices. It appeared that brain activity pertaining to attentional processing was reduced when

listening to one's own voice relative to others' voices. Xu and colleagues (2013) showed that speakers were significantly better at identifying their own voice than voices of other, familiar speakers. This effect was especially strong in difficult listening conditions, where the recordings were filtered so that only frequencies above 2000Hz were available (Xu et al. 2013). This benefit was explained by richer and more stable self-representations that could result from higher auditory familiarity with one's own voice relative to others' voices, and from strong associations with motor/articulatory representations. In line with this account, Shuster (1998) found that children with a phonological disorder accepted recordings of their own words as correctly produced more frequently than words with similar errors that were uttered by other children. That is, children recognized their own errors less often than others' errors. Shuster (1998) argued that this lower awareness of self-produced errors resulted from the experience the children had with their own, erroneous productions that eventually might have led to imprecise representations and hence imprecise perception and articulation models (see also Strömbergsson, Wengelin, & House, 2014).

"Unusual" pronunciation is common not only in the field of phonological impairments but also in the field of second language learning, where learners often retain a foreign accent. L2 acquisition models such as the Speech Learning Model (SLM, Flege, 1995) propose that foreign accents can be explained through sound similarities between the learners' first and second language (in addition to developmental and social factors)[9]. The general idea is that L2 sounds are perceived through a native-language (L1) filter. That is, L2 sounds that do not occur in a learner's first language are preferentially interpreted as the closest native category. In other words, second language learners are (at least initially) bad perceivers and this has repercussions on their productions in the second language where similar assimilation processes take place.

The perception and production of an unfamiliar sound contrast in the L2 is especially difficult if both sounds are mapped onto a single native category. The English vowel contrast /ɛ/ – /æ/, for instance, challenges learners whose first language is Dutch or German (e.g., Bohn & Flege, 1992; Broersma, 2012; Eger & Reinisch, 2019b; Escudero, Hayes-Harb, & Mitterer, 2008; Llompart & Reinisch, 2017; for learners of other L1's see e.g.,

[9] Similar predictions based on phonetic (dis-)similarities between first and second language sound contrasts are made by the Perceptual Learning Model (PAM-L2, Best & Tyler, 2007) and the Native Language Magnet Model (NLM-e, Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008).

Flege, Bohn, & Jang, 1997; Ingram & Park, 1997; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005). For learners of both languages, one of the sounds, /ɛ/ (as in *bet*), is familiar as it sounds similar to their native unrounded front open-mid vowel. Using their native substitute for this sound usually results in acceptable productions as perceived by native English listeners. It is hence an "easy" sound to acquire. However, the more open /æ/ (as it occurs in *bat*) does not occur in either German or Dutch and is difficult to acquire, therefore it is frequently substituted with the nearest L1 sound /ɛ/. To native listeners the learners' intended /æ/ hence sounds accented or simply "wrong" because to them it does not sound like the intended category. Other problematic contrasts for learners of English are voicing contrasts in word-final obstruents. Although German does have voicing contrasts in word-initial and medial obstruents, in word-final position all obstruents are devoiced and this process of devoicing is transferred to L2 English (Smith, Hayes-Harb, Bruss, & Harker, 2009; for learners of other L1's see e.g., Broersma, 2005, 2010; Cebrian, 2000; Cho & McQueen, 2006; Flege, Munro, & Skelton, 1992; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Weber, Broersma, & Aoyagi, 2011; Xie & Fowler, 2013).

Critically, whatever sound contrast should be produced, for native and nonnative speakers, there is usually more than one way to signal the contrast by using a combination of different cues. The word-final voicing contrast in English stops, for instance, is signaled by the duration of the preceding vowel, closure duration, duration of the burst/aspiration as well as voicing during the closure (Barry, 1979; Port & Dalby, 1982; Wright, 2004). One central observation when looking at such difficult but lexically relevant contrasts in second language learning is that even if a difficult contrast is not neutralized by the L2 learner, this does not necessarily mean that it is maintained in a native-like manner, neither in production nor in perception. Instead, the cues that L2 learners produce to indicate a difficult L2 contrast might be less differentiated and/or different from native cues. Similar differences for the use of cues can also be found in perception. In many cases, learners transfer their habitual use of cues from the native language to the target language, even if these cues are only secondary or irrelevant in the L2 (e.g., Bohn, 1995; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Levy & Law II, 2010; McAllister, Flege, & Piske, 2002). Importantly, it has been demonstrated that patterns of transfer may be even learner-specific (Schertz, Cho, Lotto, & Warner, 2015, 2016; Smith & Hayes-Harb, 2011) and/or differ between perception and production (Eger & Bohn, 2015; Kartushina & Frauenfelder, 2014; Kassaian, 2011; Schertz et al. 2015).

Given this influence of the learners' first language sound patterns on their accent, it has been suggested that other nonnative speakers of a target language are often as good or even better at perceiving accented speech than native speakers of the target language (Imai, Flege, & Walley, 2003; Munro, Derwing, & Morton, 2006), specifically when listeners and speakers share the same native language (Bent & Bradlow, 2003). This advantage has been termed the matched interlanguage intelligibility benefit and has been argued to result from shared knowledge about the phonetics of the first language (Bent & Bradlow, 2003, see also, Hayes-Harb et al. 2008). It may also result from incorrect perception such that L2 learners who perceive L2 speech through the perceptual filter of their native language don't suffer from mismatches if the accented production (at least partially) matches the L1 filter. This possibility seems likely given that the interlanguage intelligibility benefit has mainly been found for learners at low levels of proficiency, that is, when low-proficiency learners listen to utterances produced by other low-proficiency speakers of the L2 (Hayes-Harb et al. 2008; Pinet, Iverson, & Huckvale, 2011; van Wijngaarden, Steeneken, & Houtgast, 2002; Xie & Fowler, 2013).

In addition to this L1-filtered perception, L2 learners, especially those who learn their L2 in a classroom in their home country, have ample exposure to accented speech, for instance, from their fellow learners and often even from the teacher. A large body of research has shown that native listeners are able to quickly adapt to or tune in to nonnative speech, such that with experience accented speech becomes easier to understand (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Reinisch & Weber, 2012; Sidaras, Alexander, & Nygaard, 2009; Witteman, Weber, & McQueen, 2013). This may also be the case for second language learners. Weber, Di Betta and McQueen (2014), for instance, showed that Dutch listeners had little difficulty in processing Italian-accented English when they were familiar with an Italian accent in their L1 Dutch. That is, listeners transferred their knowledge on a specific foreign accent from their first language to a second language (see also Reinisch, Weber, & Mitterer, 2013). In addition to accent adaptation for other speakers, second language learners may also adapt to their own personal accent, that is, their very own specific L2 production patterns.

In terms of second language sound representations the suggested adaptation can be illustrated using a belief-updating model (Kleinschmidt & Jaeger, 2015). The model suggests that listeners perceive sounds as probabilistically falling into one or the other category. For our purposes of describing L2 listening we refer to them as the global sound distributions that represent the L2 categories that learners perceive through their L1 filter

and that are formed by or abstracted from native and nonnative tokens that listeners encountered. In addition, the model suggests that listeners track consistencies within the speech signal for a given situation. If consistencies are found for a given situation, the global distributions are adapted. That is, listeners are likely to have distributions for accented (L2) speech that they frequently encounter (see e.g., Cutler, 2015, for suggestions about "accented" representations). These adapted distributions can be reapplied in perception when the situation or speaker is recognized again, hence facilitating perception. Further adaptation to the situation or speaker would then proceed from this model.

Importantly, when listening to accented speech, L2 learners have experience not only with their fellow learners' accented speech but also with their own accented productions. That is, every time they speak their L2, they hear themselves speak and receive proprioceptive feedback from their (accented) productions (i.e., articulation patterns; Abbs, Gracco, & Cole, 1984; Guenther, 2006; Postma, 2000). Critically, in this way they may be even more familiar with their own speech patterns than with others' accented speech. If through this reinforcement, learners adapted to their own personal accent, they then may show a benefit when recognizing words in their own voice in the L2 relative to other learners' voices. An advantage when perceiving one's own voice could be in principle the same as the above mentioned matched interlanguage benefit, differing in that the adaptation is even more particular and applies to one's own voice rather than to an entire language-specific accent. Our aim is to experimentally test whether learners are indeed better at understanding words produced by themselves than words produced by other, equally proficient L2 learners.

Experiment 1 tested this by means of a word-reconstruction (or word-identification) task with minimal word pairs containing difficult sound contrasts for German learners of English. The contrasts were the vowel contrast /ɛ/ – /æ/ and the voicing distinction in word-final fricatives and stops. /ɛ/ and the voiceless obstruents have corresponding sounds in German and are therefore supposedly "easy" to produce. /æ/ and word-finally voiced obstruents do not occur in German and are hence difficult for Germans to produce. These sounds tend to be produced as the respective other sounds of the contrasts (Bohn & Flege, 1992; Smith et al. 2009).

German learners of English were recorded producing these words and two months later were invited back for a perception experiment. There they were presented the words one at a time and asked to decide which word of the minimal pair was intended. Critically, they were presented with words they had produced themselves as well as productions of

other learners from the sample. Speakers were grouped separately for each sound contrast such that they were matched in the type and magnitude of cues they produced. If a self-benefit was found here, that is, if learners understood/reconstructed the words better when presented with their own voice, this may suggest that familiarity with their own productions (through adaptation or stored representations) affected L2 processing.

As for the magnitude of the benefit of understanding one's own voice it would be reasonable to assume that the benefit may be larger for poor learners. This second hypothesis is motivated by previous findings that low-proficient speakers were better at understanding strongly accented speech than high-proficient and native listeners (Hayes-Harb et al. 2008). If exposure to one's own productions played a role, poor listeners might benefit even more from experience with their own accent, since their productions are overall more difficult to understand. Reversely, a self-benefit might be not so large for better learners, since the acoustic cues that they produce – even though possibly not native-like – might already be sufficient for good perception. With regard to the words within a minimal pair, the self-benefit may be larger for words with the difficult sounds, that is, the ones that do not occur in the learners' L1, as each learner may use different sets of cues to keep these sounds apart from the easy ones (i.e., the ones that are present in the L1). The role of the magnitude of produced acoustic difference between the "easy" and "difficult" sounds of the contrasts across proficiency groups will be further explored in Experiment 2.

3.2 EXPERIMENT 1

3.2.1 Method

3.2.1.1 Participants

Twenty-four female[10] students at the University of Munich participated for pay. They were native speakers of German and reported no history of speech, language, or hearing problems. The mean age was 22.4 ranging from 19 to 28. All speakers had learned English at school starting at an average age of 10.0 years and following classes for an average of 8.2 years. None of them had lived in an English-speaking country for longer than 6 months. Since on German television films and series are dubbed into German, exposure to native speakers of English was likely limited. All participants took part in two sessions: one for

[10] Since participants had to listen to their own voice and other unfamiliar voices, we decided to restrict participants to one gender in order to avoid large acoustic differences between their own and other's voices.

the recordings, and one for the perception experiment a few weeks later. In addition, they filled in a language background questionnaire with special focus on their history of learning English. The production data is shown in Figures II.B.1 to II.B.3 in the Appendix.

3.2.1.2 Materials

Thirty-one English minimal word pairs that differed in sound contrasts that have been shown to cause problems for German learners were selected (Bohn & Flege, 1992; Smith et al. 2009). Eleven minimal pairs contained the vowel contrast /ɛ/ – /æ/, seven pairs a word-final voicing contrast in fricatives and thirteen pairs a word-final voicing contrast in stops. Within each pair, one word was considered to contain an easy sound for the learners (i.e., /ɛ/, and the voiceless sounds in word final position). The other word contained a difficult sound (/æ/ and the word final voiced sounds). The labels easy and difficult were based on whether or not the critical sounds occur in the German sound inventory (German does not have the vowel /æ/) and in the given word position (German word-final phonologically voiced obstruents are canonically produced as devoiced). An additional 22 words were selected to serve as fillers in the recording session. Some of them contained other difficult sounds that do not occur in German (e.g., /θ/ or /w/) to distract participants from the main purpose of the study. Words are listed in Appendix II.A.1.

3.2.1.3 Recordings

For the recordings all words were randomly assigned to one of ten semantically neutral carrier sentences such as *The next word is...* (see Appendix II.A.2). Target words always occurred in sentence final position. The assignment of sentences to words and the order of words within the recording session were randomized separately for each participant with the restriction that the words of a minimal pair could not follow one another. Each word was repeated twice for a total of 160 sentences[11].

Participants received all instructions in English and were asked to read out the entire sentence at a comfortable pace. The sentences including the target words were presented one by one on a screen, and a small light signaled when to start speaking. The recordings were made in a soundproof recording room using a diaphragm microphone (Neumann Microphone, type TLM 103) and the software Speechrecorder (Draxler & Jänsch, 2004), which stored each sentence as a separate wav file on a computer. After the session, participants had to review a list of all target words (in randomized order) and mark those

[11] The words *bet*, *bat*, *bed* and *bad* were used for the stop voicing contrast as well as for the vowel contrast, but each word was recorded only twice.

words that seemed unknown or unfamiliar to them. The whole session lasted approximately 50 minutes.

3.2.1.4 Acoustic analyses

Several acoustic measures were taken using Praat (Version 5.4.08, Boersma & Weenink, 2015) for each sound contrast in order to group participants by production pattern for the perception experiment. The grouping of participants was done separately for each type of sound contrast, based on how well speakers differentiated the two words of the minimal pairs in production. We will refer to this grouping according to production patterns by the term "proficiency", note however, that this measure is not based on how native listeners judged the individual productions. Rather, with the term proficiency we refer to the acoustically measured magnitude of the produced difference between the critical sounds in the words of the minimal pairs (for native speakers' goodness ratings of a subset of the present productions see Eger & Reinisch, 2019b). We use the term "proficiency" rather than "production accuracy" or other terms highlighting production measures since in the perception experiments "proficiency" will refer to the *listeners'* proficiency (i.e., listeners' own production accuracy) rather than the quality of the sound material that they were listening to (for details see below).

The magnitude of the speakers' produced contrasts was assessed relative to the other speakers in the experiment such that differences between speakers within a group were minimized. Figures II.B.1 to II.B.3 in the Appendix show the main acoustic measures for each type of contrast for each of the eventually-formed proficiency groups. For the vowels, these cues were the between-category differences in the first two formants and duration. For the word-final fricatives, these were the duration of the preceding vowel and the duration of the fricative (combined as vowel duration divided by fricative duration), and the voiced portion of the fricative. For the word-final stops, the duration of the aspiration, the duration of the preceding vowel and the voiced portion of the closure were taken into account. These are the cues that are reported in the literature to be the most important ones for native speakers of English (see e.g., Deterding, 1997; Hillenbrand, Getty, Clark, & Wheeler, 1995, for the vowels; e.g., Broersma, 2010; Wright, 2004, for the fricatives; e.g., Barry, 1979; Smith et al. 2009, for the stops). The cues to the production of each contrast were weighted in the order named above. Specifically, we looked how *differently* the acoustic measures of the two categories were produced. A good contrast was defined as a large difference between the means of these measures within minimal pairs. "Good" also

indicated that categories were discrete as reflected in smaller standard deviations for each measure and thus less overlap between the words of the minimal pairs. Looking at these measures of how large a difference the twenty-four participants had produced for each of the three sound contrasts it was decided that a split into three "proficiency" groups (A= best, B = middle, C = worst) of eight participants would best capture the main differences. The assignment to groups was done separately for each contrast and followed this procedure: First, for each type of contrast separately, the eight best speakers with the clearest contrasts were assigned to the most proficient group A. Next, the eight speakers with the smallest/worst contrasts were assigned to group C. The remaining eight speakers were then assigned to group B.

In order to reduce the number of unfamiliar voices per participant presented in the perception experiment, within each of the three proficiency groups two subgroups were formed such that each group contained only four instead of eight voices (i.e., each listener was presented with only three unfamiliar voices per contrast). In this way, it was ensured that a sufficient amount of data could be collected for the analysis for tokens in one's own voice (see also Design below). The subgroups were formed such that speakers in one sub-group of four were not only similar in the overall amount of the produced difference for each contrast, but also according to which cues they had produced the largest difference in. This was especially important for speakers who had produced better contrasts using multiple cues since the goal was to match the productions of the participants' own voices with the presented others' voices as closely as possible. For instance, one speaker in group A may use duration of the preceding vowel to indicate a phonologically voiced fricative in *rise* in contrast to *rice*, with a longer vowel in the former, whereas another speaker from group A may produce the word-final voiced fricative with a long period of voicing in the fricative rather than a longer preceding vowel. However, for the statistical comparisons between the proficiency groups, only the three overall groups were considered, since differences between the subgroups were rather small and a comparison between all six groups appeared no more informative than a comparison between the three overall groups. Notably proficiency had to be used as a grouped variable rather than a continuous measure because the manipulation of Voice (listening to one's own voice vs. others' voices) had to be compared within sets of participants that listened to each other.

3.2.1.5 Design

For the perception experiment, the words of the minimal pairs were spliced out of the carrier sentences to be presented in isolation. Of the 31 recorded pairs, three were excluded due to incorrect production of sounds other than the target sounds. Additionally, the word pair *latter-letter* was excluded because several participants indicated in the questionnaire that they did not know the meaning of the word "latter". The final word set consisted of the remaining 50 words (27[12] pairs, see Appendix II.A.1).

The stimulus set of the experiment was prepared separately for each participant for each of the three sound contrasts. For each contrast, participants were presented with their own productions and those of three other speakers (henceforth other voices) that had been selected to match this participant in terms of use of cues for this contrast ("proficiency"; see above). That is, overall the stimulus set for each participant consisted of a 25% of own productions. The other voices were assigned separately for each type of contrast to maximally match the production patterns such as to isolate as far as possible the effect of voice over the types and magnitude of acoustic cues used to produce the contrasts. The total number of other unfamiliar voices in the experiment varied between 5 and 9 (though mostly 7 or 8) per participant. This was because a specific unfamiliar voice could occur in one, two, or even all three sets of contrasts. Within each contrast each participant always heard three other voices. That is, although one's own voice was heard more frequently than any other single, unfamiliar voice in the experiment, overall it was heard less often than unfamiliar voices (i.e., 25%). Using groups of four learners per sound contrast appeared the best way to divide up our set of participants to tightly control production patterns between the own and other voices for each sound contrast. At the same time, it allowed us to collect enough data points for own-voice trials which would have been substantially lower had we compared every participant to everyone else or lowered the number of own-voice trials to the number of trials for each individual voice in the group.

For each voice, the two recorded repetitions of each word were presented twice each (i.e., for a total of 4 repetitions per voice per word). All words were presented once before they were repeated. The experiment consisted of a total of 864 trials (27 pairs x 2 words x 4 speakers x 2 spoken repetitions x 2 blocks) of which 216 tokens were in the participants' own voice and 648 in an unfamiliar voice.

[12] The words *bet*, *bat*, *bed* and *bad* were used for the stop voicing contrast as well as for the vowel contrast, therefore the word set consisted of 27 pairs (25 plus additional 2) but only 50 single words.

3.2.1.6 Procedure

Participants returned for the perception experiment approximately six weeks after they were recorded. They were informed about the procedure by means of written instructions. It was noted (although not emphasized) that among several unfamiliar voices they would hear themselves. Instructions were written in English as to set participants into an English language mode without influencing their perception by talking to them with a specific accent. Participants were seated in a sound-proof cabin in front of a laptop computer. On each trial, participants saw the two words of the minimal pair written on the left and right side of the computer screen. After 400 ms they were presented one of the words over headphones at a comfortable listening level. Their task was to indicate by button press which of the words was intended by the speaker. They pressed the 1-key on the computer keyboard if they thought the speaker intended the word on the left, and the 0-key for the word on the right. The material was presented in randomized order and the position (left or right) of the two response alternatives was counterbalanced according to correct answer so that participants were not biased towards left or right position. The experiment was implemented in Psychopy2 (Version 1.83.01; Peirce, 2007), and took approximately 50 minutes to complete. Every 70 trials participants were allowed to take a self-paced break. After the experiment participants were asked whether they had recognized their own productions throughout the experiment which most of them confirmed.

3.2.2 Results

Listeners' responses were categorized into correct and incorrect answers depending on whether they chose the intended word (i.e., the word the speaker intended to produce) or the other member of the minimal pair. Correct vs. incorrect (coded as 1 and 0 respectively) was used as the dichotomous dependent variable in a series of linear mixed-effects models with a logistic linking function (Jaeger, 2008). The models were implemented in R (Version 3.3.0, R Core Team, 2017) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The random-effects structure included random intercepts for participant and word with random slopes for fixed factors that were manipulated within participants and items respectively. Note that the models with a full random effects structure (Barr, Levy, Scheepers, & Tily, 2013) did not always converge. In this case, it was determined via model comparisons using log-likelihood ratio tests whether simpler models would fit the data just as well. The best fitting model with the largest random effects structure that converged will be reported.

Since our main hypothesis was that listeners would be better at recognizing the intended word when hearing their own voice than other speakers' voices, the main variable of interest was Voice (i.e., whether they heard themselves or not). This variable was contrast coded to 0.5 = self, -0.5 = other. Note that "other" was collapsed across the different other (not-self) voices since within each proficiency group (i.e., groups of listeners that had produced similar cues to the sound contrasts) each voice contributed to "self" as well as "other" trials.

An overall model on all data with only Voice as fixed factor revealed a significant effect of Voice ($b_{(Voice)}$=0.24, SE=0.07, z=3.41, p<.001; $b_{(Intercept)}$=1.18, SE=0.17, z=6.87, p<.001) showing that overall in the experiment more correct answers were given if the stimulus was in the participant's own voice than someone else's voice ($M_{(self)}$ = 75.1 % correct, SD = 0.43; $M_{(other)}$ = 71.9 %, SD = 0.45) of the same proficiency level (i.e., because voices were matched on production patterns).

To test whether the effect of hearing one's own voice only emerged over the course of the experiment because listeners heard their own voice more often than any single other voice (since proficiency was matched within sound contrast), Trial number and its interaction with Voice were added as fixed factors to the model. Trial Number was centered and scaled from -1 to 1. As in the overall model, there was a significant effect of Voice ($b_{(Voice)}$=0.24, SE=0.07, z=3.42, p<.001; $b_{(Intercept)}$=1.18, SE=0.17, z=6.86, p<.001). However, neither Trial Number ($b_{(TrialsNumber)}$=0.05, SE=0.04, z=1.49, p=.14) nor the interaction between Trial Number and Voice were significant ($b_{(Voice:TrialNumber)}$=-0.03, SE=0.07, z=-0.42, p=.68). This indicates that the effect of Voice did not change over the course of the experiment. Trial Number was therefore not included in the subsequent analyses. The effect of Voice was then tested in relation to a number of other independent variables: sound Contrast (vowels, fricatives, stops), Proficiency (grouped into A, B, and C) and Sound Type (easy, difficult).

3.2.2.1 Sound Contrast

To test whether the effect of Voice held for all three types of sound contrasts (i.e., minimal pairs differing in the vowel contrast /ɛ/–/æ/, the final voicing contrast in fricatives or stops) a model was fitted with Voice, sound Contrast, and their interaction as fixed factors. Sound Contrast was coded as a factor with three levels with the vowel contrast mapped onto the intercept. Results showed that participants performed better overall for words from the stop contrast ($M_{(stop)}$ correct = 81.1%, SD = 0.39) than for words from the vowel contrast

($M_{(vowel)}$ correct = 67.2%, SD = 0.47; $b_{(Intercept\_vowel)}$=0.47, SE=0.17, z=2.68, p<.01; $b_{(Contrast\_stop)}$=1.76, SE=0.20, z=8.85, p<.001) but overall performance for the fricative contrast did not differ from the vowel contrast ($M_{(fricative)}$ correct = 66.5 %, SD = 0.47; $b_{(Contrast\_fricative)}$=0.46, SE=0.26, z=1.77, p=.08). Critically, for the vowel contrast that had been mapped onto the intercept there was a significant effect of Voice ($b_{(voice)}$=0.20, SE=0.08, z=2.46, p<.05) with better performance if participants heard their own voice. There was no interaction between Voice and sound Contrast for either other level of this factor suggesting that the magnitude of the effect of Voice that was found for the vowels was not different for the stop or fricative contrasts (fricatives: b=0.08, SE=0.12, z=0.65, p=.52; stops: b=0.22, SE=0.21, z=1.05, p=.29). Given that the effect of Voice was not modulated by sound Contrast, this factor was not included in any further analyses and data were collapsed across contrasts. Note that this did not affect the grouping of participants, which had been conducted for each sound contrast separately, and was coded in the variable Proficiency.

3.2.2.2 Proficiency

To test the hypothesis that participants with a lower proficiency in English would benefit more from hearing their own voice, a model was fitted with Voice, Proficiency and their interaction as fixed factors. Proficiency was taken to refer to the groups A, B, and C that listeners were assigned to according to their productions, that is, the magnitude with which each of the contrasts had been produced. Consequently, when we henceforth refer to the "high-proficiency group", or "group A", we refer to those participants who had produced a well-differentiated contrast between the words of the minimal pairs. Participants from the "low-proficiency group", group C, had produced the smallest contrast according to the acoustic measures. Participants from group B performed at an intermediate level. As described in the Methods section in more detail, this grouping was done separately for each sound contrast. Note that since in this experiment participants were listeners, we will refer to this factor as also "listener proficiency". While in Experiment 1, participants only heard stimuli from their own proficiency group, and listener and speaker proficiency (i.e., how well the stimuli had been produced) are confounded, this distinction will be relevant for Experiment 2. The factor Proficiency was coded as numeric with A = 0.5, B = 0, and C = -0.5. With this coding, the grand mean is mapped onto the intercept and effects can be interpreted as main effects.

As can be observed in Figure 3.1, the benefit of hearing one's own voice was present for all three of our proficiency groups. This was confirmed by statistical analyses. There was a significant effect of Voice ($b_{(Voice)}$=0.24, SE=0.08, z=3.04, p<.01; $b_{(Intercept)}$=1.21, SE=0.15, z=8.00, p<.001). As expected, Proficiency had a significant effect such that the higher the proficiency the more correct responses were given ($b_{(Proficiency)}$=0.55, SE=0.13, z=4.22, p<.001). The interaction between Voice and Proficiency was not significant suggesting that the effect of Voice was not different between proficiency groups ($b_{(Voice:Proficiency)}$=-0.17, SE=0.14, z=-1.19, p=.23).

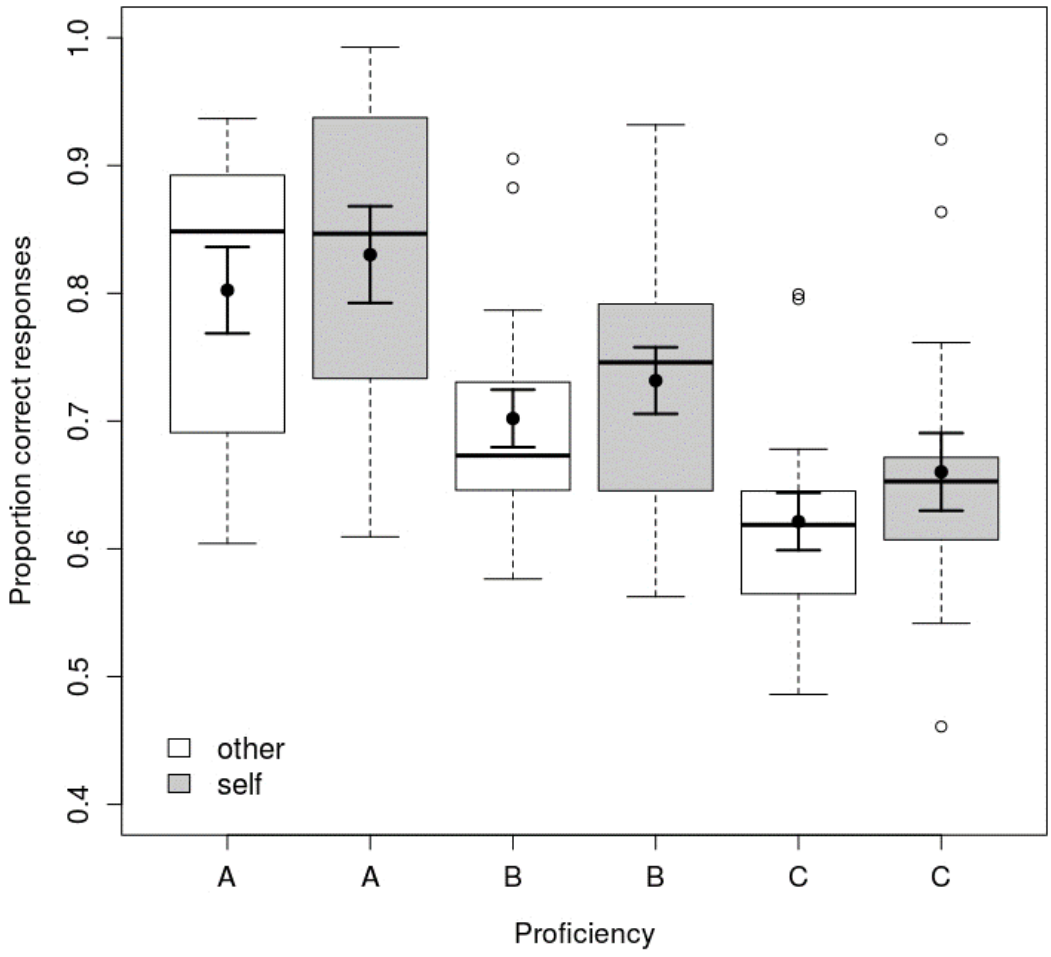

Figure 3.1: Proportion of correct responses in Experiment 1 for three proficiency groups, shown for participants' own voice (self) and others' voices, averaged over contrast. Data points are shown aggregated over repetitions and words. Chance performance is at 0.5.

3.2.2.3 Sound Type

Sound Type refers to whether the intended word of the minimal pair contained the sound that listeners know from their L1 (i.e., /ɛ/ and the voiceless obstruents as "easy" sounds; contrast coded as 0.5) or not (i.e., /æ/ and the voiced obstruents as "difficult" sounds; coded as -0.5). An interaction with Voice would indicate that the effect found for Voice (with more correct responses for one's own voice) differed according to whether the word contained an easy or a difficult sound. Sound Type was added to the model including Proficiency since even though Proficiency did not interact with Voice in the analysis reported above, the identification of difficult sounds may especially challenge participants with lower proficiency (Hayes-Harb et al. 2008; Pinet et al. 2011; van Wijngaarden et al. 2002; Xie & Fowler, 2013). Note that since all dependent variables were contrast coded, again the grand mean is mapped onto the intercept and effects can be interpreted as main effects.

Results are shown in Figure 3.2 and Table 3.1. Again, there was a main effect of Voice (more correct responses for one's own voice), and Proficiency (more correct responses the higher the proficiency; that is, the better participants had produced the contrasts themselves). The effect of Sound Type was not significant. However, these effects were modulated by a two-way interaction between Sound Type and Proficiency as well as the three-way interaction between all factors. These interactions are illustrated in Figures 3.2 and 3.3. While there were more correct responses for words containing the difficult sound category for participants of proficiency group A, in proficiency group C more words containing the easy sound were recognized correctly. The three-way interaction suggests that this modulation of Sound Type by Proficiency had repercussions on the Voice effect. That is, the effect of Voice differed between proficiency groups when considering easy and difficult categories separately.

Follow-up analyses on the three-way interaction testing each Sound Type separately revealed that in words containing the easy sound, there was a significant effect of Voice with more correct answers when hearing one's own voice ($b_{(Voice)}$=0.28, SE=0.13, z=2.19, p<.05; $b_{(Intercept)}$=1.32, SE=0.18, z=7.33, p<.001). Proficiency and the interaction between Proficiency and Voice just failed to reach significance ($b_{(Proficiency)}$=0.28, SE=0.15, z=1.89, p=.06; $b_{(Voice:Proficiency)}$=0.38, SE=0.21, z=1.78, p=.08). The benefit when hearing one's own voice for easy words hence did not differ between proficiency groups. If anything there was a slight tendency for the self-benefit to become larger for higher proficiency groups (i.e., from proficiency group C to A, Figure 3.3 left panel).

For words with the difficult sounds, an effect of Proficiency emerged with more correct responses the higher the proficiency group ($b_{(Proficiency)}$=0.96, SE=0.18, z=5.35, p<.001; $b_{(Intercept)}$=1.19, SE=0.19, z=6.17, p<.001) and an interaction between Proficiency and Voice ($b_{(Voice:Proficiency)}$=-0.61, SE=0.22, z=-2.77, p<.01). The effect of Voice was not significant ($b_{(Voice)}$=0.15, SE=0.13, z=1.12, p=.26). That is, when looking at the difficult words, the self-benefit appears larger the lower the proficiency (i.e., in proficiency group C; see Figure 3.3 right panel).

Table 3.1: Results of the mixed-effects model fitted with all factors (Voice, Proficiency, and Sound Type) in Experiment 1.

Fixed effect	*b*	SE	*z*	*p*
Intercept	1.23	0.16	7.89	<.001
Voice	0.22	0.08	2.89	<.01
Sound Type	-0.26	0.15	-1.76	=.08
Proficiency	0.56	0.12	4.80	<.001
Voice:Sound Type	0.13	0.21	0.61	=.54
Voice:Proficiency	-0.15	0.15	-0.99	=.32
Sound Type:Proficiency	-0.77	0.21	-3.61	<.001
Voice:Sound Type:Proficiency	1.00	0.30	3.29	<.001

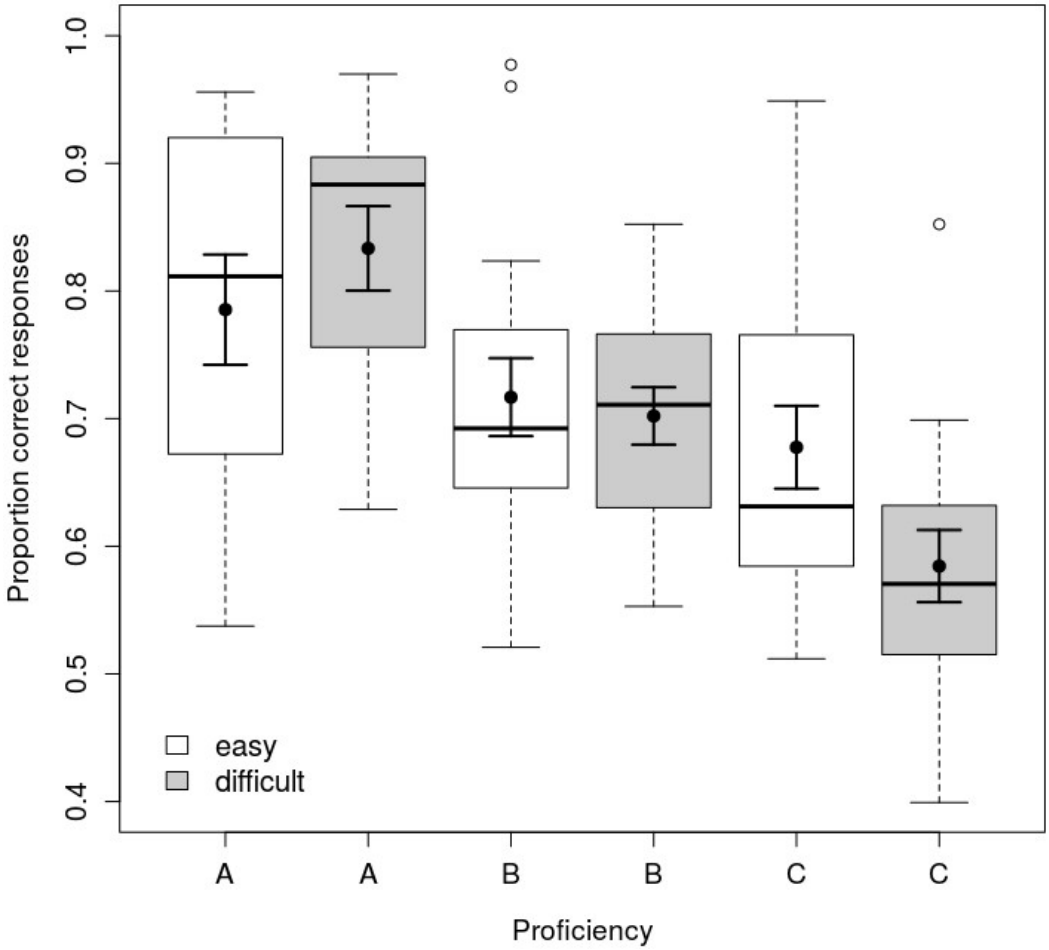

Figure 3.2: Proportion of correct responses in Experiment 1 for the three proficiency groups, shown for easy and difficult sounds, averaged over contrast. Data points are shown aggregated over repetitions and words. Chance performance is at 0.5.

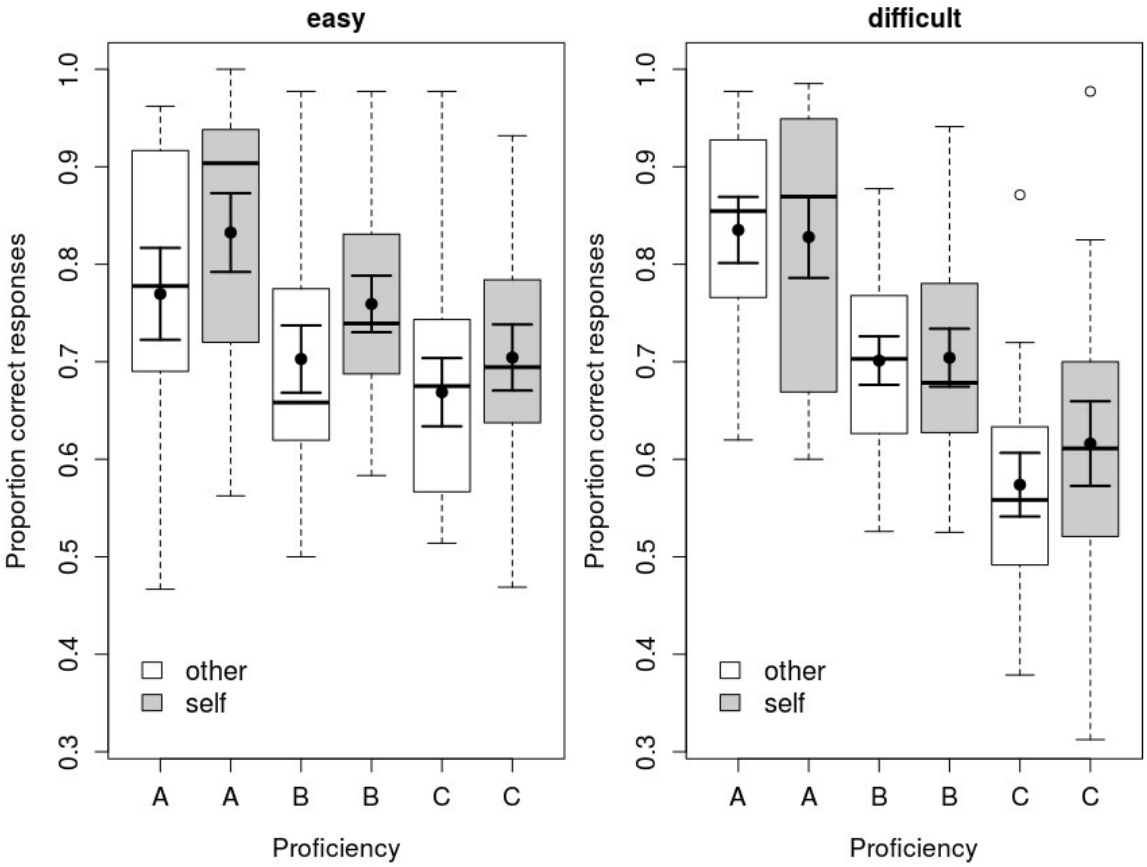

Figure 3.3: Illustration of the three-way interaction between Sound Type (easy, difficult), Proficiency (A, B, C), and Voice (self, other) in Experiment 1. Proportion of correct responses are averaged over contrast. Data points are shown aggregated over repetitions and words. Chance performance is at 0.5.

3.2.3 Discussion

Experiment 1 demonstrated that second language learners are better at recognizing L2 words containing sounds from a difficult L2 sound contrast if they had produced the words themselves than when listening to other learners of the same L1 that used similar acoustic cues to produce the L2 words. Although the overall ease of identifying the intended word differed between sound contrasts, the effect of Voice did not differ. This suggests that the effect of Voice was not specific to one of the contrasts but may constitute a more general effect.

In an overall analysis including Voice and Proficiency as factors, the effect of Voice did not differ between proficiency groups (i.e., the variable that refers to listeners' own production skills). That is, as could be expected, there was a main effect of Proficiency

such that the higher the proficiency of the learners, and hence the clearer the sound contrasts had been produced, the more correct responses were given. However, the hypothesis that the effect of hearing one's own voice may be larger for poor learners (i.e., poor producers), could not be confirmed in an all or nothing fashion (i.e., lack of an interaction in this analysis).

Interestingly, the effect of Voice was modulated in a three-way interaction with Proficiency and Sound Type. The latter defines whether the specific word of the minimal pair contains the sound that is similar to the learner's first language (i.e., “easy”) or the sound that is exclusive of the L2 (i.e., “difficult”). Looking first at the results for words with the easy sound, that is, the L2 sound that is similar to the corresponding sound in the learners' L1, there was a main effect of voice (i.e., better recognition of words produced in one's own voice) and this self-benefit was not different between the three proficiency groups (though see Figure 3.3, left panel, for the tendency that for the easy sounds the self-benefit was somewhat larger the higher the proficiency). Considering words containing the difficult sounds, in contrast, the self-benefit was larger the lower the proficiency of the learners (i.e., interaction between Voice and Proficiency in the follow-up analyses; see Figure 3.3, right panel). This confirms our suggestion that the self-benefit may be largest for poor learners, at least for difficult sounds.

The reason for the modulation of the effect of Voice by Sound Type together with Proficiency is likely to be found in the acoustics that the learners produced or failed to produce for the difficult sounds. As can be seen from the summary of the production data (Figures II.B.1 to II.B.3 in the Appendix), the learners that were assigned to Group A produced large contrasts between the words of the minimal pairs, specifically by better cuing the difficult sounds. The low-proficiency group, in contrast, produced rather small differences between the words in the minimal pairs (henceforth “poor” productions). Therefore, for the difficult sounds poor learners could benefit more from knowing their own patterns, since they had to rely on overall much smaller and/or less reliable cues to identify the difficult sounds. Note that despite these modulations of the effect of Voice, listeners from all proficiency groups benefitted from listening to their own productions in most conditions.

Why would listeners benefit from listening to their own productions? As discussed in the introduction, one reason may be the frequent exposure to one’s own accent. That is, whenever one speaks, one also hears one’s own productions, hence, unless an L2 learner is only passively exposed to the L2, their own voice is likely one of those heard most often

also in the second language. Therefore, listeners are likely highly familiar with the relation between their own productions (or production strategies) and the acoustic consequences of these strategies for perceiving these difficult sound contrasts. Note that this does not necessarily mean that our listeners would remember how they read the specific word list during recordings six weeks prior to the perception experiment. Rather they may rely on adapted representations of accented sounds as their L2 targets.

In order to compare the perception of listeners' own vs. other voices, participants in Experiment 1 were matched according to how clearly and with what types of cues they had produced the contrasts. Consequently, participants who had produced better contrasts were also presented with "rich" tokens, and low-proficiency speakers were presented with "poor" material. The next question then is what would happen if poor learners had larger/more cues available and the good learners smaller/fewer cues. In order to investigate the influence of the availability of cues and the learner's proficiency, a second experiment was designed. It aimed at showing whether in Experiment 1 low-proficiency participants performed worse than high-proficiency participants only because they were presented with "poor" tokens or because they are also less capable of picking up acoustic cues. Moreover, if being a better producer helped L2 perception in general then the high-proficiency learners should outperform low-proficiency learners regardless of the magnitude of available cues.

3.3 EXPERIMENT 2

The aim of this experiment was to test the interaction between the availability of acoustic cues in the L2 speech signal and how well participants had produced the cues themselves (i.e., what we labeled "proficiency"). Specifically, we asked whether poor learners would benefit in perception from receiving more differentiated acoustic cues to the difficult sound contrasts. To address this question, those participants of Experiment 1 who differentiated the contrasts most and those who differentiated them least in production were invited back for a second perception experiment (i.e., groups A and C). They performed the same word identification task as before, but this time they were presented with stimuli of the opposite proficiency group. That is, participants who had been assigned to the high-proficiency group based on their productions were now presented with the "poor" tokens (i.e., those produced with only few and small acoustic cues) and participants who were assigned to the low-proficiency group were presented the "rich" tokens (i.e., those produced with various and clearer cues). Since the availability of cues may play a crucial role for word recognition, we expected that, when presented with "rich" productions, low-proficiency learners may

perform better than when presented with poor productions (as in Experiment 1). However, the crucial question was whether despite this expected improvement, they would reach the level of the high-proficiency listeners found in Experiment 1. If not, then the availability of cues may play a role in L2 perception but the ability to use these cues may be related to the learners' own production abilities[13]. As concerns the high-proficiency learners, the question is whether they may be better at recognizing the words than the low-proficiency learners regardless of the quality of the stimuli (i.e., when presented the rich and poor tokens). This would suggest that the ability to pick up available cues in perception is strongly related to the learner's ability to produce these cues. An additional comparison of group C listeners' perception of poor vs. rich material with the perception of the stimuli that they had produced themselves (i.e., in Experiment 1) will show how the effects of listener proficiency, hearing one's own voice, and availability of acoustic cues relate to each other.

3.3.1 Method

3.3.1.1 Participants

A subset of participants from Experiment 1 was invited to return for a third session. To test for differences in production and perception skills, the subset of those participants was selected who differentiated the contrasts best and those who differentiated them least in production. Specifically, we selected those participants who had been assigned to group A or C in Experiment 1 for at least two of the three sound contrasts and who were not in the opposite proficiency group for the third contrast (for example, a high-proficiency participant was either only in A-groups for all contrasts, or she was assigned to A for the stops and the vowels, and B for the fricatives). Seven participants from the high-proficiency group returned as did eight participants from the low-proficiency group. The experiment was run 4 to 11 weeks (mean=6.7 weeks) after the first perception experiment.

3.3.1.2 Materials

Since in Experiment 1 the sets of stimuli varied for each participant (due to the individual combination of the participant's own and several other voices), two sets of four participants from Experiment 1 were chosen from good speakers and two from poor speakers. The two sets from good speakers consisted exclusively of tokens produced by speakers from the

[13] Note that we refer to production abilities here, since we grouped participants by production measures. However, we cannot determine the direction of causality. Our high-proficient participants could be good producers because they are good perceivers rather than the other way around. This will be further considered in the General Discussion.

high-proficiency group A. The two sets from poor speakers consisted of only tokens that had been produced by speakers from group C. Two sets per condition were chosen in order to use a representative sample of voices similar to Experiment 1. Each set contained between 8 and 10 voices across the three sound contrasts. Each participant was presented with one of these sets. Note that since in this experiment participants listened to stimuli of speakers from the "opposite" proficiency group, all voices were unfamiliar to them.

3.3.1.3 Design and Procedure

The design and procedure were the same as in Experiment 1. The participants received the same instructions but were told that this time they would not hear their own voice, but only "new", unfamiliar voices. They heard the words in isolation and had to decide which word of the minimal pair had been produced.

3.3.1.4 Analyses

Again, listeners' responses were categorized into correct and incorrect responses depending on whether they chose the intended word or the other word of the minimal pair. Responses were coded as 1 = correct and 0 = incorrect, respectively, and used as the dichotomous dependent variable in a set of linear mixed-effects models with a logistic linking function (Jaeger, 2008). As for Experiment 1, the random-effects structure included random intercepts for participant and word with random slopes for fixed factors that were manipulated within participants and items. The best fitting model with the largest random effects structure that converged will be reported.

To compare high- and low-proficiency learners when presented with both, the "rich" and "poor" tokens, subsets of data from Experiment 1 were included in the present analyses. Note that our variable proficiency was based on production. However, since this grouping was done separately for each sound contrast, some of the participants invited back for Experiment 2 had been assigned to group B for one of the contrasts. In order to restrict our analyses to data from group A or C, all trials from contrasts for which a given participant had been assigned to group B in Experiment 1 were excluded from the analyses. For example, a given participant in Experiment 2 was assigned to group C for the fricatives and the stops in Experiment 1, because she produced the contrasts with very few and small acoustic cues, but she was assigned to group B for the vowel contrast. After participating in Experiment 2 as a listener of the overall proficiency group C, her responses to words from the vowel contrast were discarded. In this way, we controlled for both the quality of the tokens and listener proficiency to be restricted to groups A and C.

For the first model, only data for "other" voices were used, since in Experiment 2 voices were necessarily others' voices. The main variables of interest were then the availability of acoustic cues (i.e., Material with the levels rich and poor, coded as 0.5 and -0.5 respectively) and Proficiency of the listener (coded as 0.5 for the participants from the high-proficiency group A, and -0.5 for participants from the low-proficiency group C; note that again proficiency refers to the production as discussed in Experiment 1). In addition, an interaction between these variables was specified. As for Experiment 1 additional analyses included the factors Contrast, to test whether results held for all three sound contrasts, and Sound Type, to test whether effects would differ for easy and difficult sounds.

3.3.2 Results

The analysis of the model with Material, Proficiency, and their interaction as fixed factors revealed a significant effect of Material ($b_{(Material)}$=1.14, SE=0.14, z=8.40, p<.001; $b_{(Intercept)}$=1.11, SE=0.12, z=9.42, p<.001) with more correct responses for the rich than the poor tokens (see Figure 3.4, in white and dark grey boxes). There was a significant effect of listener Proficiency ($b_{(Proficiency)}$=0.54, SE=0.15, z=3.64, p<.001) with more correct responses for listeners from the highly-proficient group. Moreover, Proficiency was involved in an interaction with Material ($b_{(Material:Proficiency)}$=0.60, SE= 0.20, z=2.95, p<.01), indicating that the effect of Material was different in the two proficiency groups. Results are shown in Figure 3.4 in white and dark grey boxes.

To follow up on the interaction, two additional analyses were run to test the effect of Proficiency within each Material set. Proficiency was contrast-coded to A=0.5 and C=-0.5, as before. The results revealed significant effects of Proficiency for both the rich and the poor material set, with more correct responses when the listener was from the high-proficiency group A (rich material: $b_{(Proficiency)}$=0.84, SE=0.23, z=3.59, p<.001; $b_{(Intercept)}$=1.69, SE=0.16, z=10.32, p<.001; poor material: $b_{(Proficiency)}$=0.25, SE=0.11, z=2.36, p<.05; $b_{(Intercept)}$=0.54, SE=0.10, z=5.27, p<.001). This together with the difference in regression weights (i.e., higher $b_{(Proficiency)}$ for the rich than poor material set) suggests that the interaction between Proficiency and Material is driven by the magnitude of the effects. That is, both high- and low-proficiency learners can benefit in word recognition from hearing rich over poor cues, but learners from the high-proficiency group benefit to a larger extent.

Since poor learners appear to benefit from rich material, the question arises as to how this effect compares to the effect of Voice (i.e., the self-benefit) found in Experiment 1. Therefore, in an additional analysis, the responses to "self"-trials from Experiment 1 were added to the dataset described above - again only for those participants who participated in both experiments. To compare poor learners' performance on rich material to all other conditions, one combined variable with six levels was included in the model instead of the previously used variables Proficiency and Material. Two of those levels defined the "self"-trials for each proficiency group (i.e., *A self*, *C self*). The other four consisted of trials in other voices, once with material from the same and once with material from the opposite proficiency group: listeners *A* hearing *poor* material, listeners *A* hearing *rich* material, listeners *C* hearing *poor* material, and listeners *C* hearing *rich* material.

In order to specifically test poor listeners' responses when presented with rich material relative to their performance when hearing their own voice (with poor cues) the level *C rich* was mapped on the Intercept. Results are presented in Table 3.2 and Figure 3.4 (*C rich* is indicated by thicker lines) and revealed significant differences between *C rich* and all other levels of the variable. That is, listeners from the low-proficiency group that were presented with rich material (*C rich*) performed significantly better than listeners from the same proficiency group when presented with their own (*C self*) or with others' productions from the poor material set (*C poor*). Moreover, those listeners were also better than listeners from the high-proficiency group when presented with poor material (*A poor*). Finally, low-proficiency listeners presented with rich material performed significantly worse than high-proficiency listeners hearing rich material, in both cases where high-proficiency listeners heard their own voice (*A self*) or other voices (*A rich*). This suggests that for low-proficiency listeners who themselves differentiate difficult contrasts only by using few and poor cues, the advantage when being presented with rich cues goes beyond and above the self-benefit. For learners who have already reached a high level in production in the L2 and produce more differentiated cues to the contrasts, the self-benefit appears to be on top of the material effect (see the effect of Voice in Experiment 1 for all proficiency groups).

Table 3.2: Results of the mixed-effects model to compare the effects of Proficiency, Material, and Voice with reference to the poor listeners and rich material (i.e., C rich mapped on the intercept; see text for details) in Experiment 2.

Fixed effect	*b*	SE	*z*	*p*
Intercept (C rich)	1.14	0.12	9.13	<.001
C self	-0.66	0.07	-9.29	<.001
C poor	-0.72	0.05	-14.42	<.001
A rich	0.86	0.13	6.53	<.001
A self	1.39	0.16	8.72	<.001
A poor	-0.50	0.13	-3.92	<.001

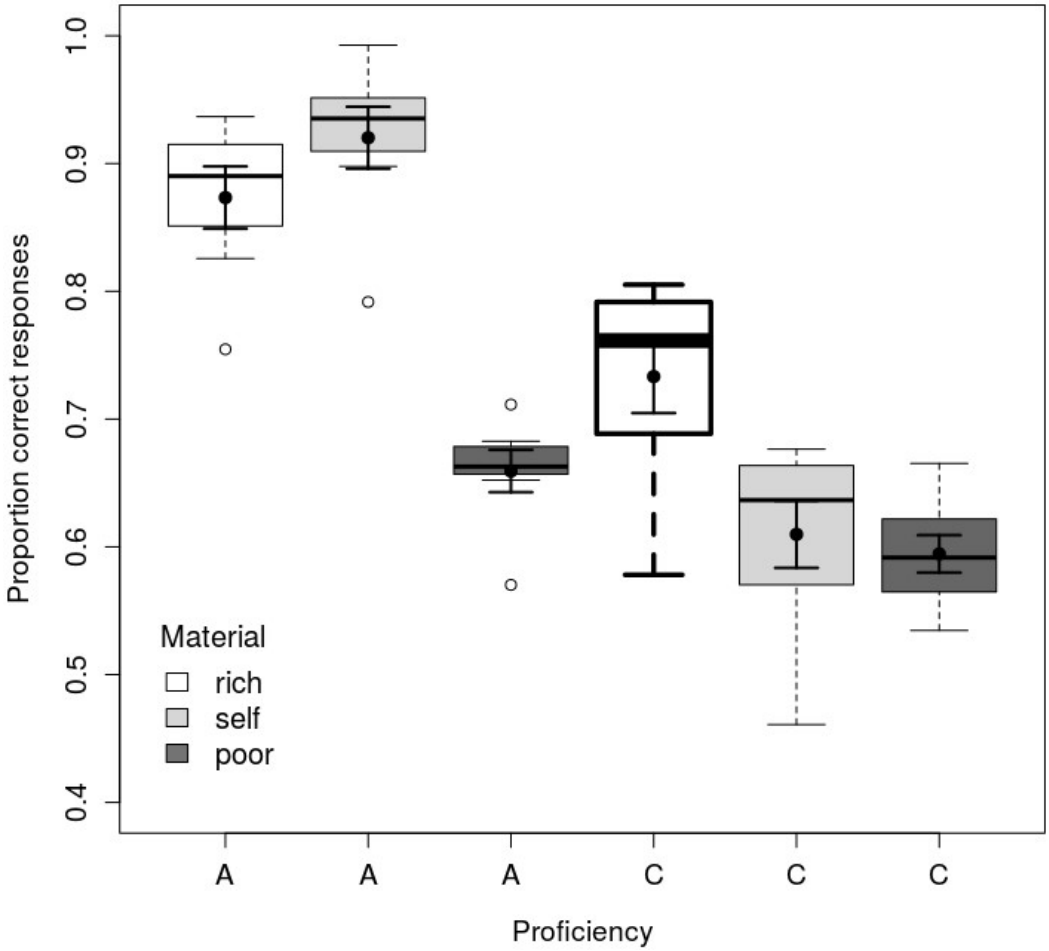

Figure 3.4: Proportion of correct responses in Experiment 2 for the two material conditions (rich and poor) and the tokens in the participants' own voices, shown for two subgroups of participants from the proficiency categories A and C, averaged over contrast. A subset of the data was added from Experiment 1 (the responses to "self"-trials and to tokens produced by others from the same proficiency group). Data points are shown aggregated over

repetitions and words. Chance performance is at 0.5. The box with thick lines refers to the condition that was mapped onto the intercept in the statistical analyses (see text for details).

3.3.2.1 Sound Contrast

Having established that the effect of Material for the poor listeners goes above and beyond the self-benefit, for ease of interpretation, the remaining analyses will focus on the effect of Material without the factor Voice. To test whether the effects of Proficiency and Material reported above differed between sound contrasts, additional analyses were run involving the factor sound Contrast with the level "vowels" mapped onto the intercept (i.e., as in Experiment 1). Effects for the other contrasts are then interpreted relative to this reference level. Results are given in Table 3.3. They suggest that all effects found in the overall analyses held for the vowels (i.e., effect of Material, Proficiency, and their interaction). Furthermore, the same effects in the other contrasts were not significantly different from the effects in the vowels (i.e., Proficiency:Contrast was not significant for either fricatives or stops). Only two significant differences were found between the vowel contrast and either fricatives or stops: First, words from the stop contrast were overall identified better than words of the vowel contrasts (effect of Contrast for Stops, $M_{(stops)}$ = 80.2 % correct, SD = 0.40; $M_{(vowels)}$ = 65.6 % correct, SD = 0.47; $M_{(fricatives)}$ = 65.8 % correct, SD = 0.47). Secondly, the interactions between Contrast and Material for the fricatives and stops indicate that the effect of Material, that is, the difference in overall correct responses between rich and poor tokens, was larger for the fricatives (21.6 % difference of correct responses; p=.06) and stops (20.3 %; p<.05) than the vowels (10.0 %; see also Table 3.3).

Table 3.3: Results of the mixed-effects model fitted with Material, Proficiency, Contrast, and their interactions in Experiment 2.

	Fixed effect	*b*	SE	*z*	*p*
Vowels	Intercept	0.42	0.16	2.62	<.01
	Material	0.73	0.19	3.94	<.001
	Proficiency	0.55	0.23	2.38	<.05
	Proficiency:Material	0.96	0.29	3.32	<.001
Fricatives	Contrast	0.45	0.27	1.65	=.10
	Material:Contrast	0.59	0.31	1.89	=.06
	Proficiency:Contrast	0.04	0.36	0.12	=.91

	Proficiency:Material:Contrast	0.11	0.48	0.23	=.82
Stops	Contrast	1.51	0.13	11.51	<.001
	Material:Contrast	0.65	0.28	2.32	<.05
	Proficiency:Contrast	-0.21	0.20	-1.08	=.28
	Proficiency:Material:Contrast	0.66	0.47	-1.40	=.16

3.3.2.2 Sound Type

The effect of Sound Type was entered to the model with Material and listener Proficiency as fixed factors. This was to test whether – as in Experiment 1 – the effects of Material and Proficiency differed between words with an easy or a difficult sound (again coded as 0.5 and -0.5, respectively). Results are given in Table 3.4 and Figure 3.5. As in the overall analysis, there was an effect of Material and an effect of listener Proficiency in the same directions as before, and the interaction between these two. Again, the interaction indicates that high-proficiency listeners benefit more from rich cues than listeners from the low-proficiency group do. Furthermore, there was no main effect of Sound Type, but one significant interaction between Sound Type and Material: As can be observed in Figure 3.5, the effect found for Material, that is, that words produced with rich cues (white boxes) are understood better than words with fewer and poorer cues (grey boxes), was larger for words containing the difficult sound category (two-way interaction between Material and Sound Type). The marginally significant three-way interaction between Sound Type, Material and Proficiency suggests that this effect was somewhat larger for the high-proficiency listeners (see also Figure 3.5).

Table 3.4: Results of the mixed-effects model fitted with Material, Proficiency and Sound Type in Experiment 2.

Fixed effect	*b*	SE	*z*	*p*
Intercept	1.14	0.12	9.56	<.001
Material	1.16	0.14	8.39	<.001
Proficiency	0.58	0.15	3.89	<.001
Sound Type	0.12	0.20	0.66	=.51
Proficiency:Material	0.67	0.20	3.26	<.01

Proficiency:Sound Type	-0.11	0.33	-0.32	=.75
Material:Sound Type	-0.43	0.20	-2.16	<.01
Proficiency:Material:Sound Type	-0.52	0.27	-1.89	=.06

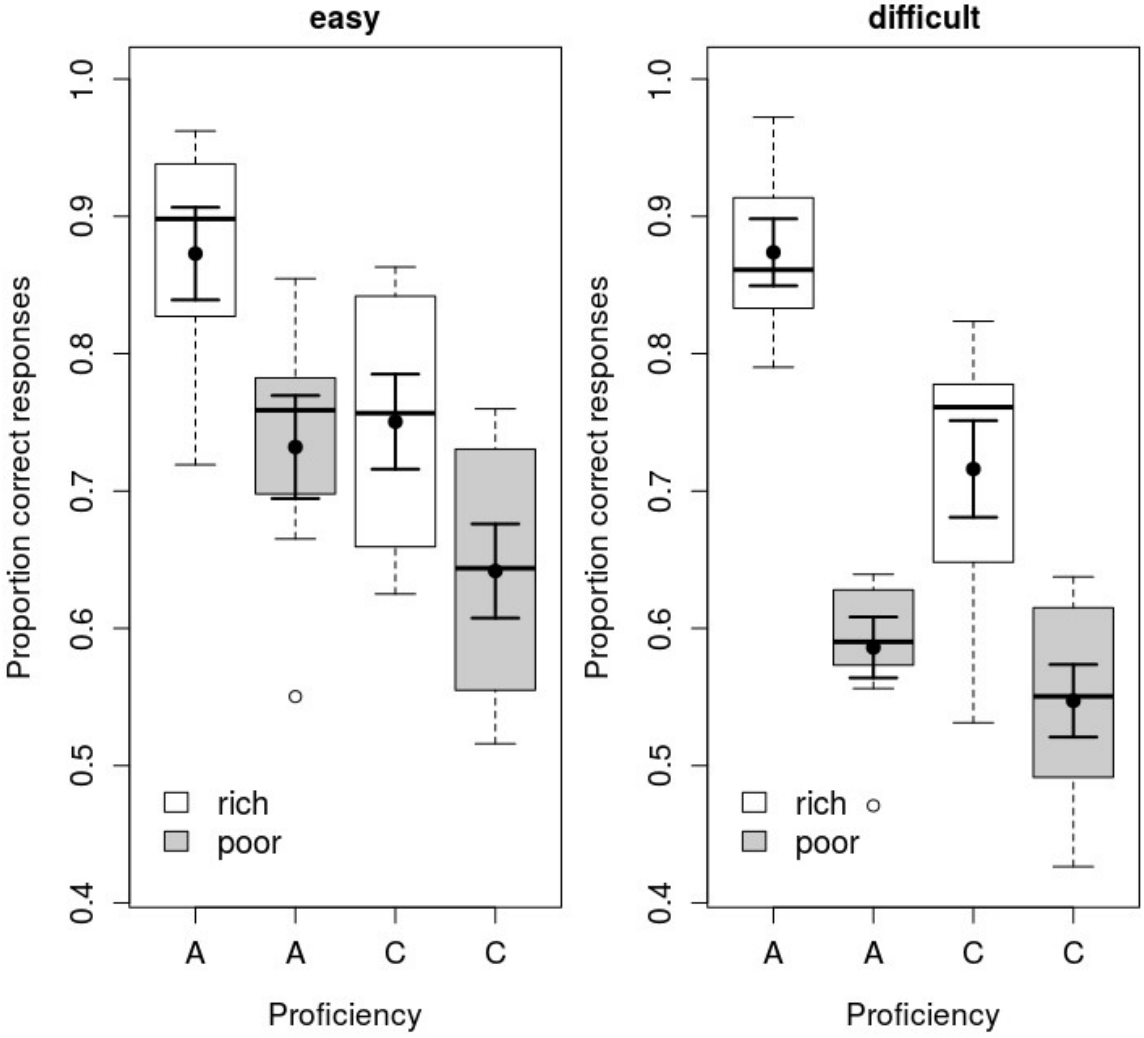

Figure 3.5: Proportion of correct responses in Experiment 2 for the two proficiency groups of participants (A and C), for poor and rich speech Material split by Sound Type (i.e., easy vs. difficult sounds). Data points are shown aggregated over repetitions and words. Chance performance is at 0.5.

3.3.3 Discussion

Experiment 2 tested to what extent the complex pattern of results found in Experiment 1 with regard to proficiency and contrast could be due to the availability of larger acoustic differences between the words of the minimal pairs for the high-proficiency group. Specifically, we asked, firstly, whether poor learners would benefit in perception if they were also presented with large acoustic differences between the words of the minimal pairs and specifically more differentiated cues on the difficult sounds (i.e., the sounds that they don't know from their L1). Secondly, we asked whether good learners would outperform poor learners regardless of the magnitude of available cues.

Results showed that indeed, poor learners benefit from "rich" speech material, that is they were better at recognizing words produced by speakers from group A than their "own" group C productions in Experiment 1. However, despite this benefit for the "rich" materials, low-proficiency learners did not reach the level of performance of the proficient listeners (see Figure 3.4). Critically, high-proficiency learners outperformed poor proficiency learners for both the rich and poor material, however, the effect was larger for the rich materials. This suggests that differentiating L2 contrasts better in production (cf. our definition of proficiency) allows learners to perceive even small cues to an L2 contrast better than poor learners, however this benefit is seen especially then, when sounds are cued in a fashion that approaches native production.

A comparison of the effect of Material with the effect of Voice (i.e., the self-benefit shown in Experiment 1) revealed that even though poor listeners benefit from hearing their own poor productions over others' poor productions, the availability of rich cues in the signal leads to even better word recognition. That is, the effect of rich material was above and beyond the self-benefit for low-proficiency learners. When the learners' productions were already clearly differentiated (i.e., in proficiency group A), the benefit when hearing one's own, good productions was on top of the effect of rich material produced by other speakers.

The benefit of being presented "rich" over "poor" productions was present for all three sound contrasts, but differed in its magnitude. It was larger for the fricative and stop contrasts than for the vowel contrast. As can be seen from the production data in Appendix II.B.1-II.B.3 the produced acoustic differences between the words of the voicing contrast in word-final stops or fricatives was larger than for the vowel contrast. Together with the results of Experiment 1 where listeners also showed overall better performance for the word-final voicing contrasts, this underlines the relevance of available acoustic cues in L2

word identification: the more differentiated the cues to a difficult sound contrast, the better L2 perception is.

The relevance of acoustic cues is further confirmed when looking at the easy and difficult sounds within each contrast (Sound Type). Words with the difficult sounds, that is, sounds that do not occur in the learners' L1, were overall harder to recognize when they were produced by poor speakers (i.e., the two-way interaction between Material and Sound Type, the grey boxes in the right panel of Figure 3.5). When the words with difficult sounds were produced by good speakers (i.e., were part of our "rich" productions, white boxes) then both listener groups showed a benefit, but with the above-mentioned difference between high- and low-proficient listeners (tentatively confirmed by the marginally significant three-way interaction between all factors).

Taken together, the results underline that, in order to recognize L2 words that differ in a difficult contrast, learners benefit most from two sources: (i) acoustic cues to be used for perception that result from a differentiated production of the minimal word pairs and (ii) having already acquired a reasonable level of proficiency, here established as good production skills. This can be especially observed in words containing a difficult sound. Even though low-proficiency learners had a larger benefit when hearing rich cues than when hearing their own, poor productions, the self-benefit helped identify the intended word, and was on top of the material effect when both sets contained rich cues (i.e., in the high-proficiency learners).

3.4 General Discussion

The present study showed that learners of a second language understand L2 words better when they were spoken in their own voice than others' voices, even when the speech material was matched according to the speakers' proficiency, that is, production patterns. The question of whether a self-benefit could be found in L2 learners was motivated by the hypothesis that due to frequent exposure to their own accented speech, learners are highly familiar with their own L2 sound patterns - above and beyond the sound patterns that are typical of their L1's accent - and that this facilitates the identification of the words they had produced themselves.

The self-perception benefit that we found in Experiment 1 did not differ between the three sound contrasts and proficiency groups. The hypothesis that the self-perception benefit would be larger for poor than proficient learners could only be confirmed in the three-way interaction of Voice, Proficiency, and Sound Type. Whereas the self-benefit did

not differ between groups when presented with words containing the "easy" sound of the contrast (i.e., the one that was similar to the participants' L1), for words with the difficult sound the self-benefit was mainly apparent in the low-proficiency speakers. This difference for words with the difficult sound could be explained by differences in the produced acoustic cues between proficiency groups. Tokens produced by the high-proficiency group contained more or better cues to identify the words within a sound contrast, specifically by producing clearer cues to the difficult sounds. This likely helped not only in self-perception but also in the identification of words produced by the other speakers of the group who were matched in their production patterns. In other words, the proficient groups could be reasonably sure to identify the difficult sounds as the intended ones by identifying the acoustic cues to the sounds even when they were produced by other learners. The self-benefit for the easy sounds in the high-proficiency group could stem from knowledge that despite the other good tokens in the experiment, German learners of English in general tend to use this sound as a substitution for the difficult sound. Hence the good learners were less confident to really hear the easy sound if it was produced by others than if it was produced by themselves. The low-proficiency participants, in contrast, had overall more trouble to identify the intended sounds since they themselves as well as the others they heard during the experiment had produced very small differences between the sounds of the contrasts, which made all words hard to identify (see also the main effect of proficiency). This suggests that the self-benefit is found relative to proficiency-matched other learners especially under difficult listening conditions.

Experiment 2 further investigated the role of the quality of the input in the perception of others' L2 productions. Being presented with rich material, that is, words that had been produced with more and clearer acoustic cues to the contrasts, enhanced word recognition in both proficiency groups compared to when hearing poor material. However, there was also an overall effect of proficiency: High-proficiency listeners outperformed learners from the low-proficiency group regardless of the availability of acoustic cues, that is, when presented with poor and when presented with rich productions. Given that "proficiency" was based on the learners' productions, this confirms L2 models suggesting that perception and production abilities are somehow linked. Moreover, the finding that high-proficiency learners outperformed low-proficiency learners, but that this difference was larger when presented with rich material, suggests that the pattern that better producers make better perceivers is more complex, and depends also on the type of input. A comparison with productions in the learners' own voices further revealed that the effect of material was

stronger than the self-benefit. Crucially, however, this was true only for learners from the low-proficiency group, who themselves had produced only small cues to the contrasts. Those learners who have already acquired more advanced production skills also benefitted from rich compared to poor material, but even more when they heard words that have been produced by themselves.

While the idea that L2 perception and production abilities are somehow linked is commonly-accepted, the exact relation and direction of causation remains yet unclear. Studies comparing perception and production abilities in L2 learners show mixed results (e.g., Flege et al. 1997; Hattori & Iverson, 2010; Kartushina & Frauenfelder, 2014; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015; Peperkamp & Bouchon, 2011; Schertz et al. 2015). Tsukada and colleagues (2005), for example, showed that Korean bilinguals who started to speak English early in life produced English vowels in a native-like fashion, but their ability to perceptually discriminate them failed to reach a native-like level (Tsukada et al. 2005). Results like these (see also, e.g., Kassaian, 2011; Kluge, Rauber, Reis & Bion, 2007; Sheldon & Strange, 1982) challenge the probably most common proposal that L2 perception leads production (Flege, 1995). Training studies also give insights into the relation between production and perception, with the specific focus on development, but again, results about cross-modal transfer of training are mixed (e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Herd, Jongman, & Sereno, 2013; Kartushina et al. 2015). However, one drawback of these previous studies is that they often used different tasks and types of material, for instance, acoustic measurements or intelligibility ratings by native listeners to determine production skills on one hand, identification and discrimination tests using native productions or synthetic stimuli to determine perception skills on the other (e.g., Flege et al. 1997; Hattori & Iverson, 2010).

Although the results of the present study cannot decide on the direction of causality between improvements in perception and production abilities either, its contribution was to compare L2 perception by groups of listeners who had produced the stimuli themselves and were grouped based on their production abilities. In this sense, the present results contribute to the understanding of L2 production and perception in that they show how learners of different proficiency levels exploit the cues in naturally produced stimuli. The Speech Learning Model (Flege, 1995) proposes that perception may lead production, in the sense that "the production of an L2 phonetic segment will typically be no more native-like than its perceptual representation" (Flege, 2003: 322). The present study investigated in more detail how difficult L2 sound contrasts are produced and perceived by focusing on

the use of relevant cues. They showed that both high and low-proficiency speakers were able to make use of available acoustic cues in perception, but the more proficient learners to a larger extent. Critically, across all proficiency groups the present study showed that listeners are better able at reconstructing words they had produced themselves than other learners, when they were matched in proficiency (i.e., magnitude and type of produced cues). Although this effect was smaller than the effect of Material or overall Proficiency (as shown in Experiment 2, Figure 3.4) it provides further insights into the learners' representations of L2 speech.

That is, in terms of modelling the perception process (i.e., access to these representations), the present results indicate that learners make use of different (sub)sets of cues including speaker-independent cues as well as fine-tuned cues typical of their own productions. In addition to adapting to other speakers whose productions are characterized by nonnative ways of differentiating difficult sounds or contrasts (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Reinisch & Weber, 2012; Reinisch et al., 2013; Witteman et al. 2013), learners might also adapt to their own productions in a second language. These language, accent, and speaker-specific adapted sound targets could be stored and accessed depending on context (Kleinschmidt & Jaeger, 2015; see Reinisch, 2016a,b, for similar suggestions concerning other speaker-specific cues). In other words, slightly different auditory targets may be accessed depending on whether one's own voice is perceived or another, unfamiliar voice. The idea that listeners adapt to their own productions receives support from the link between production and perception such that experience with one's own productions differs from experience with others' primarily by the combined sensorimotor and auditory feedback the learner receives immediately during speaking (Guenther, 2006; Perkell, Guenther, Lane, Matthies, Stockmann, Tiede, & Zandipour, 2004). This coupling may contribute to stronger adaptation to one's own voice than to others and may be part of the reason why the self-benefit goes even beyond the mere interlanguage intelligibility benefit that should apply to all participants in the current study, when listening to other learners of a similar proficiency.

If familiarity with one's own specific production patterns was a critical factor in our self-perception benefit, the question arises as to how strong it would be relative to benefits of other familiar voices. As simple as it seems to test this issue, a well-controlled empirical study is hard to design as it would be unclear how familiar, for example, pairs of friends would be to listening to each other in their second language - provided that their overall L2 proficiency was similar as well. However, the present comparison to others' voices that

were closely matched in production proficiency and types of cues produced provides first insights into the contribution of one's own voice in processing L2 speech. One's own voice is special since every time a person speaks not only acoustic but also proprioceptive feedback is experienced (Guenther, 2006; Perkell et al. 2004). Critically, changes in a speaker's own production patterns can cause changes in perception of the relevant sound contrasts (Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Shiller, Sato, Gracco, & Baum, 2009). Given this special role of one's own voice, the experimental design was matched as much as possible a natural situation, in which one person is confronted with multiple other, unfamiliar voices, yet hearing one's own voice more than any single other. That is, one's own voice may always be the most familiar.

One prediction that falls out of our account on own-voice sound representations is that listeners must recognize their own voice for this effect to occur. Schuerman et al. (2015; see also Schuerman, 2017) provide tentative evidence for this. When listeners were asked to transcribe noise-vocoded words that originally had been spoken by themselves vs. a speaker whose voice represented the voice of an "average speaker" (i.e., voice characteristics were most similar to all voices used in the experiment) participants did not show a self-benefit. However, in this study most listeners did not recognize their own voice. In another study where listeners transcribed short sentences produced by themselves or others in speech-shaped background noise, a self-benefit did appear (Schuerman, 2017). In that study about half of the participants reported that they did recognize their own voice. In the present study, exclusively natural stimuli were used to investigate differences between the perception of words produced by participants themselves or others. Participants were informed that they may hear their own voice and importantly, they also reported to recognize themselves during the experiment. The present study is thus in line with the suggestion that in normal listening conditions where listeners are likely to recognize their own productions, they show enhanced perception abilities for their own productions.

A final question to our account is how the self-benefit found in the present study may be related to the observation that learners have difficulties improving their foreign accent. We speculate that better recognition of one's own productions and lower awareness of potential errors may be two sides of the same coin though findings may depend on the task. A case in point is Shuster (1998) who tested the identification of speech errors by children with a phonological disorder. There the children were worse at identifying their own erroneous productions as incorrect than judging other children's incorrectly produced sounds. In the present study, the task was not to report the correctness of the productions

but to understand the words. We hypothesized that if a self-benefit can be found, this suggests that learners have adapted to their own accented (i.e., nonnative) production patterns. The familiarity with one's own “errors” (that is, own productions that are not well differentiated) may appear as an advantage in reconstructing the intended word. We speculate that as the other side of the coin this benefit may be a drawback, since learners have fewer difficulties to understand their own than fellow learners' productions, which may be one reason why the need for improvement may not be obvious.

Summing up the present study, we found that learners of all proficiency groups are able to perceptually discern difficult second-language sound contrasts, given the availability of sufficient acoustic cues marking the sounds. Listeners who are better producers themselves show an advantage in exploiting cues to second language contrasts over poor producers especially if the cues are strong. Even though clear acoustic cues contribute most to understanding difficult L2 contrasts, listeners are better at recognizing second language words produced by themselves compared to when produced by an unfamiliar speaker using similar production patterns. We hypothesized that due to frequent exposure to their own accented speech, learners are highly familiar with their own L2 sound patterns - above and beyond the sound patterns that are typical of their L1's accent - and that this facilitates the identification of the words they had produced themselves. Future research will have to show how adaptation to one’s own speech patterns and already acquired production skills in a L2 relate to the ease or difficulty to improve one’s accent in a foreign language.

4 THE ROLE OF EXPLICIT KNOWLEDGE IN THE ACQUISITION OF TWO NOVEL SOUNDS

Abstract

The present study investigated Italian learners' production and perception of German /h/ and /ʔ/ - two sounds that lack obvious linguistic counterparts in Italian. Critically, of these sounds only /h/ is explicitly known to learners from instruction and orthography. We therefore asked whether this awareness would lead to better acquisition of /h/ than /ʔ/, and whether any differences would depend on the explicitness of the task. In production, learners performed accurately in about 70% of the cases, with errors including sound deletions and substitutions. In spoken-word recognition, learners were hindered by sound deletions, but not by substitutions, although they were able to differentiate the sounds in an explicit goodness rating task. Overall, acquisition of /ʔ/ was similar to /h/, despite lack of awareness for this sound. The results suggest that learners have established one combined "glottal category" to which both sounds map in speech processing, while they may be better implemented in production.

A modified version of this chapter has been submitted for publication as:

Eger, N. A., Mitterer, H., & Reinisch, E. Learning a new sound pair in a second language: Italian learners and German glottal consonants.

4.1 INTRODUCTION

Speaking and listening in a second language (L2) that has been learned after a native language (L1) is already in place is a challenge for many learners, specifically with regard to the second language sound inventory. A wide range of empirical studies has demonstrated that those L2 sounds are especially difficult to learn that are similar, but not identical to a category in the learners' L1 (e.g., Best & Tyler, 2007; Bohn & Flege, 1992; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Escudero, Hayes-Harb, & Mitterer, 2008; Gluszek, Newheiser, & Dovidio, 2011; Hattori & Iverson, 2008; Ingram & Park, 1997; Ingvalson, Holt, & Mc Clelland, 2012; Llompart & Reinisch, 2017). In that case, learners often perceive the novel L2 sound as similar or identical to the acoustically and articulatory closest category in their native language. This leads to less accurate and slower recognition of words including these sounds (Broersma & Cutler, 2008; Pallier, Colomé, & Sebastián-Gallés, 2001; Weber & Cutler, 2004). This process of mapping L2 sounds onto the L1 is also found in L2 word production, such that learners produce L2 sounds similar to the closest native sound category even if this comes at the cost of reducing or failing to maintain critical L2 contrasts (e.g., Cebrian, 2000; Eger & Reinisch, 2019a; Llompart & Reinisch, 2017; Smith, Hayes-Harb, Bruss, & Harker, 2009).

But not all nonnative sounds that challenge L2 speakers have close counterparts in the learner's L1. One case in point is the glottal fricative /h/, which is a notorious example of a "difficult" sound for many learners, for instance, French or Italian learners of English or German. In their native languages, there is no obvious counterpart that they would typically use instead. Despite this bad reputation and despite the well-known problem of the sound /h/, there is surprisingly little work on how /h/ is acquired by L2 learners of L1 languages that lack this sound. It is possible that this is due to the special position that /h/ occupies in languages in which the sound exists and that are relatively often acquired as an L2 (such as English and German). Since in these languages /h/ tends to occupy a hermit position as it does not obviously contrast with any similar sound in these languages, prominent models of L2 sound acquisition (e.g., PAM-L2: Best & Tyler, 2007; SLM: Flege, 2003; NLM-e: Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008) would predict that this sound should be relatively easy to acquire, especially as its articulatory gesture is not particularly complex.

However, recent work on German sheds a different light on whether /h/ contrasts with any similar sound. Mitterer and Reinisch (2015) showed that native speakers of German treated /h/ and the glottal stop very similarly for lexical access, even though,

traditionally, the glottal stop is viewed as epenthetic segment that is not part of lexical representations. Phonological treatments of German (e.g., Wiese, 1996) assume that German allows vowel-initial syllables, which are then produced with an epenthetic glottal stop (e.g., /afə/ → [ʔafə], *ape*), while /h/ is a phoneme that is lexically specified (e.g., /ha:fən/ → [ha:fən], *harbor*). This predicts that deletion of an initial glottal stop and /h/ should have different effects in perception, with stronger deletion costs for the lexically specified /h/. In contrast with this prediction, deletion costs for native speakers of German were similar. These results then suggest that the two segments may form a perceptually and articulatory relatively similar contrast, giving raise to the question how L2 listeners not familiar with these sounds from their L1 deal with it.

The prediction that /h/ may be problematic for learners is partly borne out by a study that investigated how French learners of German produce word-initial /h/ (e.g., Zimmerer & Trouvain, 2015). Based on anecdotal evidence, one hypothesis was that French learners may simply delete the glottal fricative when speaking German. In contrast to that, analyses of productions in different tasks (reading single sentences, repeating sentences, and reading a text passage) revealed that learners produced /h/ in word-initial position in almost 70% of the cases. Only in very few cases they completely omitted it. In most cases in which /h/ was not realized, they produced it as a glottal stop, a sound that can occur as a stress marker in French (Malécot, 1975). These observations imply that producing the glottal stop, which also shares some articulatory properties with /h/, may be a preferred alternative for learners when /h/ is not realized successfully, rather than simply omitting it. Unfortunately, the study by Zimmerer and Trouvain did not investigate how word-initial /ʔ/ was produced by French learners, probably because the glottal stop is, in phonological theory, not seen as integral for word recognition as it turns out to be (Mitterer & Reinisch, 2015). Another study investigated phonological processing of /h/ in francophone Canadians using event related brain potentials (White, Titone, Genesee, & Steinhauer, 2017). This study used stimuli in which /h/ was either present or absent in single-word utterances. It was found that auditory discrimination abilities (as measured by the MisMatchNegativity) were good in native as well as nonnative listeners independent of proficiency. However, with regard to word recognition only native speakers and high-proficiency learners showed sensitivity to the presence vs. absence of /h/. This was measured with an N400 to pseudowords such as 'urricane (for hurricane) and havocado (for avocado). This shows that the ability to discriminate a sound does not mean that it is used for word recognition. However, similarly

to the study by Zimmerer and Trouvain discussed above, this study did not look at the processing of glottal stop.

In the present paper, we hence investigate how the two glottal sounds /h/ and /ʔ/ are acquired by Italian learners of L2 German. Neither /h/ nor /ʔ/ are phonemes in Italian[14] (Krämer, 2009) but are used to mark stress or in paralinguistic functions. The glottal fricative can only be used for instance in laughing or sighing. Notably, there are a few Italian words that are written with an initial <h>, such as *ho* ("I have"), but there it is mute. Italian also does not have wide-spread glottal marking of vowel-initial words. Vowel-initial words following voiced sounds (like other vowels or nasals) are usually realized with an "uninterrupted" transition from the preceding sound, resulting in a hiatus in the case of two vowels across word boundaries. In conversational speech, speakers also realize two adjacent vowels as a diphthong or delete one of the segments (Bertinetto & Loporcaro, 2005; Magno Caldognetto, Zmarich, & Ferrero, 1997). In hyper-articulated speech, a glottal stop may be inserted (Bertinetto & Loporcaro, 2005). Moreover, glottalization can be used as a phrase boundary marker (Stevens, Hajek, & Absalom, 2002) to indicate an open word-final syllable when it is stressed, especially in the Tuscan dialect (van Santen & D'Imperio, 1999). Interestingly, van Santen and D'Imperio (1999) also observed breathy parts, that is, parts with acoustic properties similar to /h/, in the Italians' productions, in the same context as glottalization. This may indicate that /h/ and /ʔ/ are used as interchangeable segments by Italians. Even though Italians may therefore have experience with these two sounds in their first language, it is unclear whether and how they would use them in production and perception in German. This seems reasonable because several studies found differences between coping with linguistic compared to non-linguistic sounds (Morse, 1972; see also dissociations between the articulation of speech sounds and movements of emotional expressions in patients with dysarthria: Ziegler, 2010).

To our knowledge, there are relatively few studies that examine the acquisition of two L2 sounds that are not similar to the sounds of the L1. One classic finding is that English learners are well able to discriminate click sounds from the Zulu language (Best,

[14] In the Tuscan dialect, the consonant /k/ is usually realized as glottal fricative (Marotta, 2008). On the website of the institute for German in Florence (https://www.deutschesinstitut.it/pronuncia-la-h/ , last viewed on 06/04/2018), the pronunciation of "h" in German words is even described as identical to the initial sound in "casa" (house) when spoken in the Tuscan dialect. Therefore, for the present study, Italians from the Tuscany were particularly not included, since they may have experience with producing and perceiving this sound, even if for a different phoneme category.

McRoberts, & Sithole, 1988). This finding suggests that contrasts that are not perceived as similar to any L1 sound should be acquired relatively easy (see also Faris, Best, & Tyler, 2016; 2018). However, as we review below, the ability to perform well on an auditory discrimination task does not necessarily mean that L2 learners will find it easy to make use of the contrast for linguistic processing in the L2. A case in point is the contrast between two studies on the acquisition of the Spanish contrast between tap and trill by American English learners. Both tap and trill are unlike the typical approximant realizations for /r/ in American English (Scarpace, 2014). A study by Rose (2000) had shown that naïve American English listeners are relatively good in discriminating these sounds in an AXB task. However, Scarpace (2014) used a task focusing on lexical processing and found that learners had difficulties correctly deciding whether a tap or trill that was embedded in a two-word sequence belonged to a given lexical item. This indicates that learners may treat these two sounds as free variants of one single category for lexical processing (Scarpace, 2014). Note, however, that in American English varieties, the tap is a frequent realization of the phoneme /t/ when it follows a stressed vowel (Patterson & Connine, 2001), whereas the trill may be known from other varieties of English (e.g. in Scottish; Hughes, Trudgill, & Watt, 2012). This study hence does not fully address how learners deal with two new L2 sounds that are distinct from all L1 sounds.

When investigating how a new L2 sound is acquired, the question arises how the sounds are acquired in perception and production. There have been different suggestions on whether perception and production of difficult L2 sounds are directly linked and how this relation may look. It may be the case, for instance, that the ability to perceive an L2 contrast is a necessary prerequisite for correctly producing it. Reversely, producing a contrast may enhance the ability to perceptually discern it. One famous theory of second language acquisition, for instance, proposes the former, namely that production skills depend on how well a L2 contrast is perceived (SLM, Flege, 2003). According to this idea, the same nonnative representations are accessed for production and perception, and that "without accurate perceptual 'targets' to guide the sensorimotor learning of L2 sounds, production of the L2 sounds will be inaccurate" (Flege, 1995: 238). However, studies investigating this relationship revealed mixed results. For instance, the learners' ability to perceptually differentiate sounds was not always correlated with their ability to produce them in a straightforward way (e.g., Kartushina & Frauenfelder, 2014; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015; Kassaian, 2011; Peperkamp & Bouchon, 2011; Schertz, Cho, Lotto, & Warner, 2015). Moreover, some studies reported that learners'

production abilities seemed to be better than their perception (Sheldon & Strange, 1982; Tsukada, Birdsong, Bialystok, Mack, Sung, & Flege, 2005). Training studies that investigated whether training of one modality can transfer to the other revealed that better perceptual skills due to perception training can lead to better production (Bradlow et al. 1997; Herd, Jongman, & Sereno, 2013) or reversely (Akahane-Yamada, McDermot, Adachi, Kawahara, & Pruitt, 1998). However, effects are often only weak (Bradlow et al. 1997) or cannot be found at all (Wong, 2013).

Comparing L2 production and perception abilities is further complicated by the finding that learners may perform relatively well in discriminating isolated phonemes or syllables, but at the same time perform rather poorly in tasks that tap into lexical processing (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Sebastián-Gallés & Díaz, 2012). This difference is also often related to different types of task. Discrimination of sounds is often tested in a more explicit way such as in an AXB-discrimination task, where the listener focuses on phonetic detail. To test lexical processing, in contrast, listeners typically have to perform more complex tasks, in which they are not explicitly prompted to listen to acoustic detail, such as the visual-world eye-tracking paradigm (Huettig, Rommers, & Meyer, 2011). This dissociation of explicit and implicit processing of second language sounds is supported by various other lines of research. For instance, developing listeners are able to make use of phoneme-like units in speech perception (McQueen, Tyler, & Cutler, 2012) but are usually not aware of such units until they learn to read (Morais, Cary, Alegria, & Bertelson, 1979). In cognitive neuroscience, Rauschecker and Scott (2009) suggested different auditory pathways, with a ventral stream for spoken-word recognition and a dorsal stream that links perception with production (see also Hickok & Poeppel, 2000). Somewhat in line with this proposal, Krieger-Redwood, Gaskell, Lindsay, and Jefferies (2013) found that explicit speech tasks are impaired by transcranial magnetic stimulation (TMS) to the motor regions, while an implicit task is not. This means that, even if a learner can make a distinction in an explicit task while focusing on the acoustic-phonetic properties, this may have different repercussions for the use of this contrast in spoken-word recognition and speech production.

Another important consideration in the context of the acquisition of a new L2 contrast is whether both sounds of the contrast are acquired equally well or one is acquired better than the other (Cutler, Weber, & Otake, 2006; Weber & Cutler, 2004). Considering the contrast /h/ versus /ʔ/, there are in fact reasons to assume that /h/ might be better acquired than /ʔ/. First of all, according to the current state of the art of German formal

phonology, /h/ is a phoneme and /ʔ/ is an epenthetic segment (Wiese, 1996). This would suggest that learners should, simply due to the nature of the input, acquire /h/ before /ʔ/. Secondly, /h/ is explicitly coded in German orthography with the letter <h>, while /ʔ/ is not. While there is a controversy of the role of orthography for spoken language processing in the L1 (Mitterer & Reinisch, 2015; Pattamadilok, Morais, Colin, & Kolinsky, 2014), there is clear evidence that orthography influences L2 acquisition (Bassetti, 2017; Escudero et al. 2008; Hayes-Harb, Nicol, & Barker, 2010; Simonchyk & Darcy, 2018). This is undoubtedly related to the fact that L2s are usually learned (and tested) to a great extent in the written modality. In the study on the tap-trill contrast mentioned above, for instance, learners performed worse at assigning the segment to the correct lexical item when it occurred at a word-boundary compared to word-internally. In word-internal position, the contrast is also present in orthography, where the trill is written as <rr> and the tap as <r>. At word-boundaries, the two sounds are distributed contrastively, with the trill occurring in word-initial position and the tap word-finally before a vowel, but both are written as <r>. There, the orthographic coding is hence less straightforward than word-internally, which may contribute to the difficulty learners have with the contrast (Scarpace, 2014). For the present study, this influence of orthography would hence predict that learners should be faring better with /h/ than with /ʔ/.

However, the literature on orthographic influences in the L1 also suggests that this may interact with the task effects described above. Task-dependent differences have also been shown for L1 speakers when the words differed according to their orthographic coding. In the study of Mitterer and Reinisch (2015) with native speakers of German, orthographic influences were tested both in an explicit and in an implicit task. The results revealed an orthographic effect only in the explicit task, not on spoken-word recognition in the implicit task. When explicitly asked how well the words were produced, deleting the orthographically coded /h/ had a stronger effect than deleting the /ʔ/. However, when the words had to be recognized in a visual-world eye-tracking paradigm, the reduction costs were similar, regardless of whether /h/ or the /ʔ/ was deleted. A further comparison with Maltese speakers ruled out the possibility that this was due to acoustic differences between /h/ and /ʔ/. In Maltese, the glottal stop is coded with the letter <q>, and Maltese listeners showed similar deletion costs when the glottal stop was deleted as German listeners. However, in the explicit task, an orthographic effect was also observed in the comparison between Maltese /ʔ/ vs. German /ʔ/, just as in the comparison German /ʔ/ vs. German /h/.

Finally, although Italian leaners are typically trained to produce /h/ during school lessons, they are unlikely to be ever made aware of the glottal stop[15]. Indeed, many native speakers of German are not aware of the glottal stop. For instance, the Maltese-German dictionary (written by a German linguist) claims that the glottal stop does not exist in German (Ohk, 2006). Therefore, it is unlikely that Italian learners are aware of /ʔ/ and will spend much effort on trying to get the glottal stop correct. There are hence strong reasons to expect that /h/ may be acquired better than the glottal stop.

The current study hence aims at investigating the acquisition of the German /h/-/ʔ/ contrast by learners with Italian as their L1. As reviewed above, the critical questions are how well the contrast is acquired in perception and production, whether there are differences between different perception tasks, whether /h/ may be acquired better, and whether there is an interaction of these two last factors, that is, whether the benefit of /h/ is task specific.

This was achieved in a series of three experiments. Experiment 1 tested how Italian learners produce German words starting with these sounds, and which type of non-target-like productions they would realize. Italian learners and a German control group were recorded producing German sentences, including /h/- and /ʔ/-initial target words. Italians were also recorded producing sentences with vowel-initial targets in Italian. This was done to make sure that, even if tested in a similar task, the L1 Italian does not afford frequent glottalization of vowel-initial words as one would predict on the basis of the phonological description of Italian (Bertinetto & Loporcaro, 2005).

Experiments 2 and 3 focused on Italian learners' perception of German /h/- and /ʔ/-initial words. In each experiment, perception was tested in two different types of tasks that tapped implicit and explicit processing of the contrast, respectively. To test learners' perception of words starting with the critical sounds during sentence comprehension (i.e., implicit processing), we used the visual-world eye-tracking paradigm (Experiments 2a and 3a respectively). Visual-world eye-tracking makes use of listeners' behavior to spontaneously fixate on visual referents to the acoustic input and the modulation of fixation patterns by the acoustic match between the acoustic input and the lexical representations that are accessed during comprehension (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; de Groot, Huettig, & Olivers, 2016; Huettig & McQueen, 2007; Spivey &

[15] While it is difficult to get hard data on this, an interview with a German native speaker teaching German in Italy for more than 20 years confirms that the glottal stop is only discussed at University level but not in high school.

Marian, 1999; Weber & Cutler, 2004; see Huettig et al. 2011, for an overview). In this way, by manipulating the acoustic input (e.g., substituting the critical sounds for one another) learners' target fixations may provide insights into how the two sounds are represented in the learners' lexicon. Critically, the task was implemented so that it did not require a focus on acoustic-phonetic detail. The different targets on the screen were always easily distinguishable (i.e., words starting in /h/ and /ʔ/ were never presented on the same screen). For the explicit judgment task (Experiments 2b and 3b), the same correctly and incorrectly produced target words were presented and the task was to judge how well the target word was pronounced. We expected that if learners perceive the difference between the two sounds, they should rate words with the incorrect, substituted segment as worse than words with the correct segment.

4.2 EXPERIMENT 1

The purpose of this experiment was to see how well Italian learners of German produce words that in their canonical form start with /h/ or /ʔ/. In order to avoid a reading task and thus being able to show a true orthographic effect in the acquisition of /h/ or /ʔ/ (cf., Bassetti, 2017), sentences had to be constructed from a series of pictures (see Figure 4.1 for an example) to elicit semi-spontaneous speech. The two main aims of this experiment were to find out whether the two sounds are indeed problematic for Italian learners and whether they produce not only sound deletions but also substitutions in both directions. Moreover, we expected that if explicit knowledge plays a role in L2 sound acquisition, we should see asymmetries between the two sounds with more correct productions for /h/ than /ʔ/.

4.2.1 Methods

4.2.1.1 Participants

Ten monolingual native speakers of German and 13 Italian learners of German participated for pay. The German speakers (four males) were aged between 18 and 28 (mean= 25.3, sd=3.4) and were all current or former students at the University of Munich. The Italian learners (three males) were aged between 20 and 36 (mean=29.5, sd=5.1) and came from various regions of Italy. Importantly, care was taken that no participant from Tuscany was included, since in this region the sound /k/ is typically realized as a glottal fricative. Speakers from this region may hence have more experience with producing and perceiving this sound than speakers from other regions.

Some of the Italian participants were in Munich as exchange students (n=6) whereas others had their permanent residence in Munich. According to self-report, the learners' proficiency was between B1 and C1 according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Council of Europe, 2011). Overall, they started learning German at a mean age of 19.8 (sd=8.1), ranging from an age of 6 to 36 years. Some learners started taking German lessons at school (n=5), whereas others learned German at University or by taking courses at language schools. Their mean age at the time when they arrived in Germany was 26.2 years (sd=5.3), with the youngest at an age of 18. Before their arrival in Germany, no one had longer-term contact to German spoken by native speakers. All participants filled out a questionnaire on self-reported usage and self-estimated proficiency in German. The questions could be answered on a seven-point scale, with 1 indicating frequent use, good skills, or weak accent, and 7 indicating infrequent use, poor skills, or strong accent, respectively. Table III.A in the Appendix reports the means and standard deviations of the length of residence, and of five values from the questionnaire: The self-reported frequency of speaking and listening German, the self-estimated speaking and comprehension skills, and the self-estimated accent in German.

4.2.1.2 Materials

Twenty German words were selected that start with the glottal fricative /h/ and 20 German words starting with the glottal stop /ʔ/. All words were picturable nouns and were selected to likely be known to Italian learners of German. In order to compare the learners' productions in German to how vowel-initial words are produced in Italian (i.e., whether they are truly vowel initial or whether epenthetic glottalization or glottal stops would be found), an additional 20 Italian nouns were selected. They all started with a vowel and were picturable. The German material was produced by both groups of participants (i.e., Germans and Italian learners of German). The Italian words were only produced by the Italians. Words are listed in Appendix III.B.1 and III.B.2.

4.2.1.3 Design

Participants were asked to produce the words in sentences that they had to construct from an array of pictures. The structure of the German sentences was always of the type [actor] [auxiliary for past tense] [number] [TARGET] [past participle of main verb], for instance, *Laura hat neun Hüte gekauft* ("Laura has bought nine hats"). An illustration of one example prompt is given in Figure 4.1. Across the experiment four names of the actors were used (Anna, Laura, Mario, Nico), each one indicated by a different cartoon character, and three

different verbs as indicated by symbolic pictures (see, buy, eat). Participants were familiarized with all pictures before the production task began. Critically, the target was always preceded by one of the numbers *neun, zehn* ("nine, ten") or *zwei, drei* ("two, three") ending in a sonorant sound (i.e., a nasal or a vowel). This context facilitated the detection of /h/ and the glottal stop (or glottalization) in the acoustic signal. Two repetitions of each target word were recorded, once after a nasal and once after a vowel, resulting in 80 recordings per participant.

The Italian words were presented only to the learners and should again be produced in context sentences. Sentences were elicited in a similar fashion as the German sentences but taking into account the typical Italian word order. Sentences were of the type *Nico ha comprato un buon olio per Laura* ("Nico has bought a good oil for Laura"), such that the target always followed an adjective ending in a sonorant (i.e., nasal or vowel), again to facilitate the detection of glottalization if any was present. The 20 words combined with both types of preceding context resulted in 40 Italian sentences.

Figure 4.1: Example prompt of a German sentence, with transcriptions of the full sentence that had to be constructed. Participants saw only the pictures.

4.2.1.4 Procedure

The procedure was the same in both languages and for all participants. The Italian learners were recorded producing the German and Italian sentences (in that order), the German control group only produced the German material. After reading instructions in German, each participant was shown all pictures of the target words on a paper printout and asked

to name each object. This allowed the experimenter to check whether the correct word was used, that is, to ensure that the learners were familiar with all the words and to avoid confusions of semantically related items such as *Kaninchen* ("rabbit") for *Hase* ("hare"). In cases in which an expression other than the intended one was used, the experimenter alerted the participant that the quested word was a different one. Only if participants could not guess the correct form by themselves, it was given by the experimenter, without emphasizing on the spelling or pronunciation. After this preparation, participants were seated in a sound-proof booth in front of a screen and were instructed to form sentences from the pictures and produce the whole sentence at a normal pace. The recordings were made using a diaphragm microphone (Neumann Microphone, type TLM 103) and Speechrecorder software (Draxler & Jänsch, 2004). When a speaker used a wrong target word or produced a hesitation or pause between the context and the target word the recording of this sentence was repeated.

4.2.1.5 Analysis

The goal of the analysis was to see whether and how often native speakers of German and especially the Italian learners realized the two target sounds /h/ and /ʔ/ appropriately. Productions were analyzed by two trained phoneticians[16]. The analysis was the same for both the German and the Italian sentences. First, all cases of multiple recordings of a given sentence were screened such that all recordings of sentences that contained pauses or hesitations, or in which a wrong target label had been used (e.g., *Kaninchen*, "rabbit" for *Hase*, "hare") were removed from the dataset. From the remaining sentences, the target words together with the preceding context words (i.e., the number-target sequences) were spliced out of the carrier sentences using automatic segmentation via WebMAUS (Kisler, Reichel, & Schiel, 2017) and Praat (Boersma & Weenink, 2015). The WebMAUS tool also captured remaining sentences that contained hesitations or pauses, which were excluded from subsequent analyses (93 of the 1840 German sentences produced by the learners and 1 of 800 tokens produced by the German participants). None of the 520 Italian sentences (40 trials x 13 Italian speakers) had to be excluded due to hesitations or mistaken words.

[16] Initially, it was attempted to use forced alignment for the analysis, as such an analysis is more reliable than human judgement. But it turned out that forced alignment often analysed productions by German native speakers as substitutions (i.e., /h/ for /ʔ/ and vice versa), even though the productions appeared canonical to native listeners when reviewed. This questioned the validity of forced alignment and led to the decision to use manual transcription. All these cases were then consistently transcribed as canonical by both transcribers.

These remaining recordings were manually annotated by two phoneticians, auditorily and by visual inspection of the signal using Praat. Annotators could choose between three result categories, which were /h/, /ʔ/, or deletion, for both sounds respectively. The criteria to annotate a glottal fricative were frication, that is, statistical noise in the signal, combined with a lower amplitude during the fricative as compared to the preceding and following vocalic context. Additionally, interruptions of the f0 contour were taken into account, as /h/ is described as a voiceless consonant that is produced with an open glottis. However, since all words occurred in a voiced context, this criterion was not obligatory for the glottal fricative. In German, /h/ can phonetically be fully voiced, despite its description as a phonologically unvoiced fricative. Similarly, the glottal stop is often not a complete stop (see also Ladefoged & Maddieson, 1996, p. 75; Mitterer, 2018) but is often produced as glottalization in the worlds' languages (including German, Kohler, 1994). Therefore, a glottal stop was not only transcribed if there was a clear stop of air flow, but also based on the properties of the f0 track when there was no clear stop. The criteria for the glottal stop then were either a visible deviation or interruption of the voicing (i.e., sharp drop in F0 or lack of F0 tracking). Additionally, periods of irregularity were taken as a criterion. These criteria were the same for both German and Italian sentences.

Only tokens in which both annotators agreed on the label were included in the analyses. This was the case in 92% of the remaining recordings for the German sentences. For the Italian sentences, annotators agreed in 99% of the cases. The productions of these sentences will be discussed after the analysis of the German sentences, which are the main focus of the present study.

4.2.2 Results

4.2.2.1 German sentences: learners vs. native speakers

Statistical analyses were conducted using linear mixed-effects models as provided by the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.4.3, R Core Team, 2017). Two types of analysis were performed, the first of which analyzed whether the target was produced canonically or not, and the second whether, for the non-canonical pronunciation, one target sound was more likely to be substituted or deleted. The dependent variable for the first analysis was coded with correct = 1 and incorrect = 0. We henceforth refer to "correct realization" when the annotation matched the respective target sound (i.e., /h/ and /ʔ/). The model was fitted with the fixed factors Target Sound (contrast-coded as /h/ = 0.5 and /ʔ/ = -0.5), the speakers' L1 (German coded as 0.5, Italian as -0.5), and their

interaction. With this coding, the grand mean is mapped onto the intercept, and effects can be interpreted as main effects. The random-effects structures included intercepts for participant and word (i.e., item) with random slopes for Target Sound over participants and L1 over items which amounts to the full random effects structure (Barr, Levy, Scheepers, & Tily, 2013).

Results are illustrated in Figure 4.2. Native speakers of German produced the sounds correctly in most of the cases, with only few deletions. The Italian learners produced the sounds correctly less often than the Germans, but overall correct in 69.8 % of the cases. This difference was confirmed by the statistical analyses, revealing a significant effect of L1 (b_{L1}=5.58, SE=1.14, z=4.88, p<.001; $b_{Intercept}$=3.98, SE=0.61, z=6.53, p<.001). The effect of Target and the Interaction with L1 failed to reach significance (b_{Target}=1.56, SE=0.99, z=1.58, p=.113; $b_{L1:Target}$=1.31, SE=1.82, z=0.72, p=.47). The latter might be surprising given that the difference between /h/ and /ʔ/ is much larger for Italian learners (9.8%) compared to native speakers (2.7%). Note, however, that the difference for German native speakers is near ceiling, were the logistic transformation (correctly) assigns more weight to smaller differences.

Looking at the types of incorrect productions (i.e., the two darker shadings in Figure 4.2) German speakers deleted both sounds in very few cases, but never replaced them with each other. Therefore, this second analysis was not informative for German speakers and was only conducted for Italian learners, who produced both types of non-canonical pronunciations, deletions and substitutions for both /h/ and /ʔ/. In order to test whether Italian learners deleted or replaced one of the two target sounds more often than the other, two additional models were fitted with data of the Italian learners, one for deletion, one for substitution. In both models Target Sound was the fixed factor, again coded with /h/ = 0.5 and /ʔ/ = -0.5. The random effect structure included random intercepts for participants and word, and random slopes for Target Sound over participants. In the first model, the dependent variable was whether a sound was deleted or not. Results revealed a significant effect of Target Sound, showing that Italian learners deleted /ʔ/ more often than /h/ (b_{Target}=-2.51, SE=0.93, z=-2.70, p<.01; $b_{Intercept}$=-3.35, SE=0.74, z=-4.55, p<.001). In the second model, the dependent variable was whether one sound was replaced with the other. This analysis did not reveal a significant effect of Target Sound, indicating that the substitution pattern did not differ significantly between the two sounds (b_{Target}=0.80, SE=0.97, z=0.83, p=.41; $b_{Intercept}$=-2.71, SE=0.35, z=-7.66, p<.001).

4.2.2.2 Italian sentences

The rightmost bar in Figure 4.2 illustrates how often the Italian speakers produced a glottal stop (or glottalization) when producing vowel-initial words in their own L1. Analyses revealed that in only 6.4% of the cases vowel-initial Italian words were produced with an initial glottal stop. Importantly, whether or not any glottal stop was used varied widely between participants. Of the 13 Italian participants, six never produced a glottal stop or glottalization in Italian, six produced glottalization one or two times (≤ 5%), and one produced glottalization in 42.5% of the cases. In the German sentences, the percentage of the correct /ʔ/-realizations for this speaker was relatively high, at about 82%. Overall, however, there was no correlation between the number of glottal-stop insertion for German versus Italian vowel-initial words. As expected, /h/ was never inserted.

In sum, while Italian learners of German produced very few Italian vowel-initial words as starting with a glottal stop, German words starting with /ʔ/ or /h/ were produced correctly in about 70% of the cases. Non-target-like productions included deletions and substitutions.

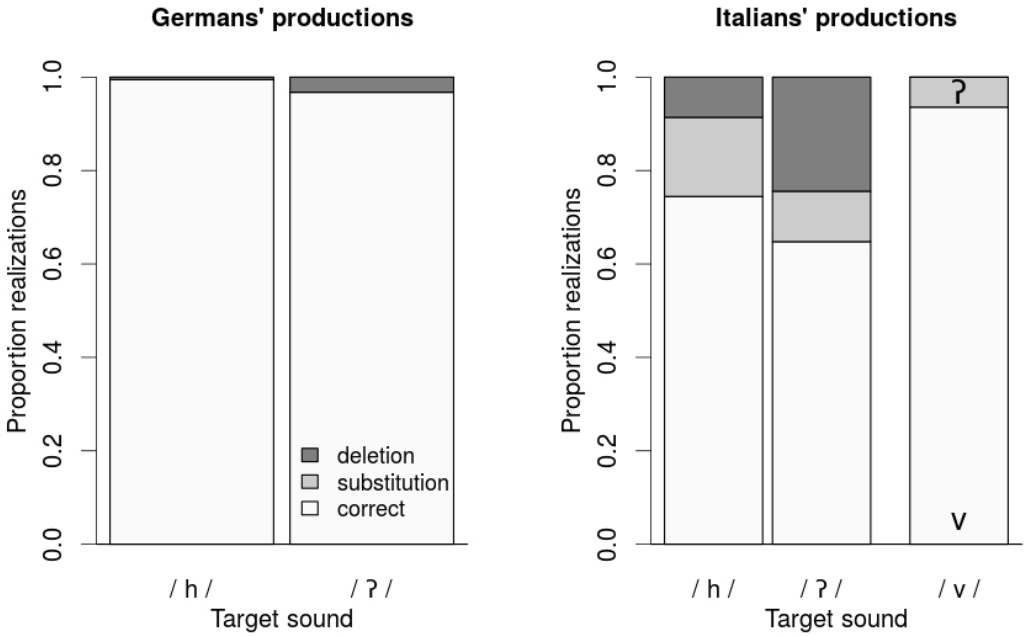

Figure 4.2: Percentage of /h/- and /ʔ/-realizations in the German sentences, shown for the German speakers (left panel) and the Italian learners (right panel). The rightmost bar in the Italians' productions shows the percentage of realizations of vowels and glottal stops in vowel-initial words in the Italian sentences by the Italian speakers.

4.2.3 Discussion

The purpose of Experiment 1 was to find out how often Italian learners of German would produce the two target sounds correctly and to see which types of non-target productions they would realize. The results demonstrated that learners indeed show problems with both sounds and produce substitutions in both directions. Previous work (Zimmerer & Trouvain, 2015) had indicated that learners with an L1 without /h/ may use the glottal stop to replace /h/. However, it was unclear whether the reverse substitution would also occur, and our data show that it does. Given that neither sound exists in the learners' L1, 70 % of correct productions for the learners indicate relatively good performance and possibly better than could be expected from anecdotal evidence.

Secondly, we asked whether /h/ would be acquired better than /ʔ/. Although the proportion of target productions did not differ significantly between /h/ and /ʔ/ among the whole group of Italian learners, the numerical difference of about 10 % more correct productions of /h/ calls for caution in the interpretation that there really is no difference. However, given the strong a-priori assumptions that /h/ should be easier, the absence of a significant effect is somewhat surprising. Apparently, learners differ strongly on how well they learn /h/ versus /ʔ/, leading to a mean difference that does not reach significance. In addition, the analyses of the learners' non-target-like productions revealed a different pattern for the two sounds: Whereas the number of substitutions did not differ significantly between /ʔ/ and /h/, the glottal stop was deleted more often than the glottal fricative.

The higher deletion rate of /ʔ/ compared to /h/ may be explained by their different status. That is, since learners are usually aware of /h/ as a problem, they may have paid special attention not to omit this sound. However, the finding that sound substitutions occur similarly often in both directions may indicate that despite learners' awareness of /h/, this sound is not clearly differentiated from /ʔ/. This leads to the question how the two sounds are used in perception. Since Experiment 1 showed that Italian learners produce substitutions similarly often in both directions, Experiment 2 compared the perception of correct productions with substituted productions (i.e., an /h/-initial word is produced with a glottal stop, and vice versa). Since performance by L2 learners typically differs between different types of tasks (Díaz et al. 2012), Experiment 2a tested the perception of the /h/-/ʔ/ contrast in an implicit task and Experiment 2b in an explicit task.

4.3 EXPERIMENT 2a

To test the perception of German /h/ and /ʔ/ in an implicit processing task we followed the example of Mitterer and Reinisch (2015) and made use of a visual-world paradigm. To focus listeners' attention on semantic properties rather than the specific target sounds, the targets were embedded in sentences in which they were predictable from the preceding sentence context. The question was to what extend Italian learners would be influenced by the substitution of /h/ and /ʔ/, and whether there would be an asymmetry between recognizing targets starting with the two sounds.

4.3.1 Method

4.3.1.1 Participants

Twenty-one monolingual native speakers of German and 20 Italian learners of German participated for pay. None had participated in the production experiment, but by specifying the same criteria for participation (between B1 and C1 according to the Common European Framework of Reference for Languages, Council of Europe, 2011), the Italian participants were matched as closely as possible to the previous sample. The German participants (five males) were aged between 18 and 34 (mean= 23.6, sd=4.7) and all of them were current or former students at the University of Munich. The Italian learners (five males) were aged between 19 and 38 (mean= 27.7, sd=5.5) and were living in Munich at the time of testing. Overall, they started learning German at a mean age of 20 (sd=7.1), ranging from 6 to 33 years. Their mean age of arrival was 25.9 years (sd=4.6), with the youngest arriving in Germany at an age of 18. Before their arrival in Germany, no one had longer-term contact to German spoken by native speakers or had lived in a German-speaking country. Some of them had their permanent residence in Munich (n=16) whereas others were exchange students. As the group in Experiment 1, the Italian learners filled in the questionnaire on their use and self-estimated proficiency in German. Overall, the mean values of the self-estimated skills in German and the self-estimated accent compared well to the overall level of the learner group in Experiment 1 (see Table III.A in the Appendix).

4.3.1.2 Materials

For the auditory and visual materials, we identified 40 /h/-initial and 40 /ʔ/-initial German picturable nouns. An additional set of 40 German words starting with various other consonants was selected to serve as fillers. For each word, a sentence was constructed such that the target was predictable from the context, and that the target was always preceded by

a word ending in a nasal. This context was chosen to facilitate the splicing procedure described below, so that the obstruents /h/ and /ʔ/ would stand out between the surrounding sonorants. Words and sentences are listed in the Appendix in Table III.B.3.

For each of the sentences, a visual display was generated that contained three pictures, which were retrieved via Google image search. On each screen, one picture referred to the target (the word in the sentence), the second was selected as a semantic competitor (i.e., also semantically fit the context preceding the target word) and the third one as a distractor that was not likely to occur in the sentence (the latter two were only presented as pictures but not recorded). For instance, for the target *Anzug* ("suit") in the sentence *Vater trägt gern den warmen Anzug* ("Father likes to wear the warm suit") the semantic competitor was *Pullover* ("sweater") and the distractor was *Nagel* ("nail"). The three words within a trial started with different consonants and were matched in estimated word frequency taking into account spoken forms (SUBTLEX-DE, Brysbaert, Buchmeier, Conrad, Jacobs, Bölte, & Böhl, 2011).

For the critical /h/- or /ʔ/-items, sentences were constructed such that the target words were, on face validity, a bit less likely than the respective semantic competitors. This was to avoid that listeners knew already from the sentence which picture was the target, which would decrease competition and hence mask potential effects of the manipulated variable (e.g., "Father likes to wear the warm …", where "suit" is slightly less likely than "sweater"). The filler sentences, by contrast, were constructed such that target words were slightly more likely than the semantic competitors, to avoid that participants "learned" that the less probable word would always be the correct one (e.g., "Leon's favorite sport is …", with the target "soccer" and the less likely semantic competitor "billiard"). The estimated probability of occurrence of the target vs. competitors in the sentences was confirmed in a pretest.

4.3.1.3 Pretest

All sentences were presented in written form in an online pretest. Each sentence was presented one after another with three dots in place of the target, and at the same time the target and the semantic competitor word were shown. Participants had to rate on a five-point scale for each of the two words how well the target and the competitor word fitted the sentence, from 1 "fits very well" to 5 "does not fit at all". Ratings from 96 students were analyzed. As intended, targets were rated as a better fit than competitors in the filler

sentences and were rated as slightly worse than competitors in the sentences including the critical /h/- and /ʔ/-items (see Appendix III.C).

The ratings were also used to assign the sentences to two lists for the eye-tracking experiment. These lists were created to split the data, so that at test, each participant would hear each word with only either the correct or the replaced segment. For each sentence, the difference between the rating of the target and the competitor was calculated, sentences were sorted according to this difference and then assigned alternately one by one to list A and list B. This was to ensure that both lists were balanced in terms of how likely the target word was compared to the semantic competitor in our two experimental conditions (see sentences in Table III.B.3 in the Appendix).

4.3.1.4 Recordings

Sentences were recorded by a female native speaker of German. Filler items were recorded once. Sentences with a critical item were recorded several times to allow cross-splicing of the recordings. That is, the sentences with a critical /h/- or /ʔ/- item were recorded at least twice in their correct form, and at least once with the initial segment replaced (i.e., glottal stop replaced by /h/ and the other way around). Three recordings were used to generate the two stimuli per item for the experiment: one of the recordings with the correct realization and one with the incorrect realization were spliced into a third recording which was used as carrier. Recordings were manipulated in Praat (Boersma & Weenink, 2015). All splicing was done at a negative-going zero-crossing at the beginning of the nasal preceding the target sound, and at a suitable place after the target sound, such that no audible artifacts resulted from cutting. If necessary, the duration of the nasal was adjusted, such that the two resulting sentences differed only in whether the target word started with a glottal stop or a glottal fricative. Additionally, in each resulting recording, the target onset was determined by visual inspection of the acoustic signal. For the critical items, the onset was defined as the end of the nasal and the start of frication for /h/, or the start of aperiodicity visible in the waveform and spectrogram (creaky voice) for /ʔ/. Target onsets of the various consonants were also marked in the filler sentences.

4.3.1.5 Design

Each participant heard 122 sentences, of which 2 were practice trials, 40 filler trials, 40 /h/-items and 40 /ʔ/-items. Half of the critical items were presented with the correct and half of them with the replaced segment, for both /h/- and /ʔ/-items. Which words were presented in the correct form and which with the replaced segment was determined by the list they

had been assigned to, based on the pretest. Lists were counterbalanced between participants. That is, each participant heard each critical item once, either with the correct or the replaced segment, but never both. The first six trials contained always four fillers and two practice trials with one correct and one replaced /h/-initial word, which were not included in the analyses. A different random order was generated for each participant, so that the order of sentences and the number of times each quadrant contained the target was balanced over participants and conditions. The experiment was implemented on SR Research Experiment Builder (Version 1.10.1630).

4.3.1.6 Procedure

Participants were seated in front of a computer screen and an Eyelink SR 1000 eye-tracker in desktop set-up was calibrated. They were instructed that they would hear German sentences over headphones and at the same time see three pictures on the screen. Their task would be to click on the picture that best matched the word in the sentence. It was explicitly phrased that they should click on the "best-matching" picture, since especially for German listeners a word spoken with the substituted segment should sound wrong. The experimenter emphasized that when the pictures appeared they should look freely all over the screen. For each trial, participants first saw a fixation cross for 600 milliseconds before the pictures appeared on the screen. After a preview of 1800 milliseconds with the pictures present, the recording started to play. As soon as the participant had clicked on one of the pictures, the next trial started automatically after 500 milliseconds, starting again with a fixation cross. The participants could take a self-paced break at three pre-defined points during the experiment after every 40 trials. The eye-tracking task took approximately 20 minutes to complete.

4.3.1.7 Analysis

Only critical items were analyzed. Of these, all words that were unknown to learners, as indicated in a questionnaire after having finished all parts of the perception experiment, were excluded from the analyses for the respective participant. One hundred and thirteen items were removed (7.1% of the Italian learners' data). From the remaining data set, those trials were also excluded in which listeners did not click on the target picture (4.1% of the remaining Italian learners' data and 0.7% of the data collected from the German group). A click was defined as correct when it was within the quadrant of the screen in which the picture corresponding to the target word was shown.

In analyzing eye-tracking data, the selection of an appropriate time window is critical. It is generally assumed that listeners need approximately 200 ms to program and launch a saccade leading to a delay of about 200 milliseconds between the onset of a spoken word and the fixation related to this input (e.g., Allopenna et al. 1998; Dahan, Magnuson, & Tanenhaus, 2001; Huettig & McQueen, 2007). However, several studies, especially with nonnative listeners report longer delays and therefore set the start of their critical time window at 300 ms (see, e.g., Cutler et al. 2006; Escudero et al. 2008; Weber & Cutler, 2004). In the present study, the time window between 300-800 milliseconds after target onset was chosen since this time window appeared to best capture listeners' reactions. This decision was based on a visual inspection of target and competitor fixations over the whole group. Independent of condition, target fixations started to diverge from fixations of competitors at about 300 milliseconds after target onset and continued to rise until about 800 milliseconds.

All results were analyzed with linear mixed-effects models with the full random effect structure (Barr et al. 2013), save for the exclusion of correlation parameters between random effects. First, one overall model was fitted including data from all participants. The dependent variable was the target fixations for the eye-tracking data, which was the logOdds-transformed proportion of fixation on the target picture over the specified time window. The model was fitted with the fixed factors L1 (German coded as 0.5, Italian as -0.5), Target Sound (/h/=0.5, /ʔ/=-0.5), Condition (correct coded as 0.5, substituted coded as -0.5), and all interactions. This way, the grand mean was mapped onto the intercept and effects can be interpreted as main effects. The random-effects structures included intercepts for participant and word (i.e., item) with random slopes for Condition and Target Sound over participant, and Condition and L1 over item (Barr et al. 2013).

4.3.2 Results

Results are reported in Table 4.1 and illustrated in Figure 4.3. Results of the overall model on fixations between 300 and 800 ms after target onset revealed a significant effect of Condition, indicating fewer looks to the target when the word was presented with the incorrect segment, and an effect of L1, which suggests fewer target fixations by the Italian learners than the German listeners. In Figure 4.3, the effect of Condition can be seen when comparing the fixation curves for the correct (light grey) and the substituted condition (dark grey), with the former rising earlier. In the German listeners (left), this is true for both /h/-items (targets with /h/ as intended initial segment, top panels) and /ʔ/-items (targets with

/ʔ/ as intended initial segment, bottom panels). The difference between the language groups can be seen comparing the fixation curves between the left (German) and the right (Italian) panels, with overall fewer target fixations by the Italian learners. L1 was involved in two significant interactions, one with Target Sound and one with Condition. The latter indicates that the effect of Condition, fewer target fixations when the words were presented with the substituted segment, differed between the two language groups. The interaction between L1 and Target Sound suggests that, although overall no difference in fixation proportions to /h/- and /ʔ/-items was found, the proportion of looks to /h/- and /ʔ/-items differed between the two language groups. All other factors and interactions did not reach significance.

To further investigate the source of the interactions, follow-up analyses with the same fixed factors as before were performed separately for the two listener groups. These analyses showed that for the German listeners, Condition was the only factor that had a significant effect ($b_{Condition}$=2.53, SE=0.33, t=7.57, p<.001; $b_{Intercept}$=5.69, SE=0.29, t=19.88, p<.001). This suggests that Germans looked less at the target picture when presented with the replaced sound, but this did not differ significantly between the target sounds (b_{Target}=0.27, SE=0.48, t=0.57, p=.57; $b_{Target:Condition}$=-1.03, SE=0.67, t=-1.54, p=.13). Analyses of the Italian learners revealed that there was no effect of Condition ($b_{Condition}$=0.37, SE=0.36, z=1.01, p=.31; $b_{Intercept}$=3.62, SE=0.27, z=13.63, p<.001) or Target (b_{Target}=-0.76, SE=0.53, z=-1.44, p=.16), and no interaction between these two ($b_{Target:Condition}$=-0.26, SE=0.73, z=-0.36, p=.72). That is, unlike German listeners, Italians did not fixate on the targets less when the word was presented with the replaced segment. Note that these follow-up analyses do not show a clear source for the interaction of L1 and Target that emerged in the overall model. This interaction may be due to a non-significant tendency that Italian listeners looked more to pictures of /ʔ/-items (fixation proportion for /ʔ/-items: 60.0%; /h/-items: 57.3%) and a non-significant tendency for more fixations on /h/-items in the German group (proportion of fixation for /h/-items: 70.3%; for /ʔ/-items: 68.6%). Since we have no hypothesis for the source of this interaction, we will not discuss it further.

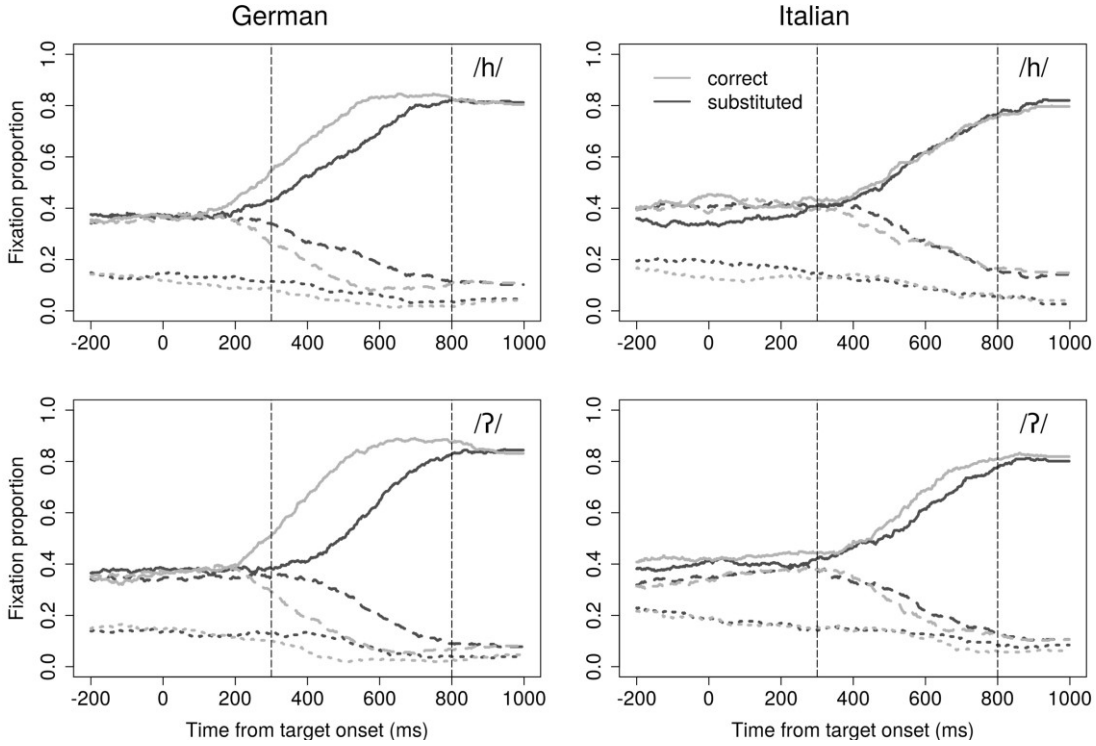

Figure 4.3: Fixation proportions on targets (solid lines), competitors (dashed lines) and distractors (dotted lines) over time (in ms; 0 is target onset) for /h/-items (top panels) and /ʔ/-items (bottom panels) in both correct (light grey lines) and substituted (dark grey lines) conditions, for native listeners of German (left) and Italian learners (right). The vertical lines indicate the time window of analyses.

Table 4.1: Results of the mixed-effects model for the logOdds-transformed fixation proportions in the critical time window between 300 and 800 ms, fitted with Condition (correct, substituted), Target (/h/, /ʔ/), L1 (German natives, Italian learners), and their interactions.

Fixed effect	*b*	SE	*t*	*p*
Intercept	4.66	0.22	20.92	<.001
Target	-0.25	0.41	-0.61	=.55
Condition	1.45	0.24	5.99	<.001
L1	2.07	0.29	7.10	<.001
Target:Condition	-0.68	0.48	-1.42	=.16
Target:L1	1.05	0.46	2.27	<.05
Condition:L1	2.15	0.47	4.59	<.001
Target:Condition:L1	-0.76	0.92	-0.83	=.41

4.3.3 Discussion

The implicit task tested to what extent Italian learners are able to make use of the /h/-/ʔ/ contrast in perception. Based on the finding that a comparable group of learners substituted both segments with one another in production, we tested to what extent Italians would be slowed down by such exchanges in perception in comparison with a control group of German native listeners. This control group performed as expected; the stimuli with exchanged segments led to slower word recognition, and this effect did not differ between the two target sounds. Italian learners were overall slower in spoken-word recognition than the native speakers, however, they were not slowed down when presented with the incorrect pronunciation. Or, from a different perspective, Italians did not have an advantage when they heard the correct pronunciation. This indicates that they cannot use the sounds in online processing of words starting with these sounds.

Overall, the performance of the Italian learners has to be considered surprisingly "bad" given the relatively strong performance on the production task. While in production, in 70% of the cases the learners produced the correct segment, an exchange of /h/ and /ʔ/ did not lead to a decrease in performance in a word-recognition task.

4.4 EXPERIMENT 2b

Since the implicit perception task in Experiment 2a showed that Italian learners of German are not affected by sound substitutions of /h/ and /ʔ/ the question arises whether they would perceive the difference between the two segments when asked explicitly. This issue was addressed in an explicit goodness rating task that all participants completed right after having finished the eye-tracking experiment. In this task, participants heard all critical target words from the eye-tracking task in minimal context and had to tell for each target in both conditions (i.e., correct vs. with the critical sounds exchanged) how well it was pronounced. To indicate the identity of each word, the recordings were presented together with the pictures of the targets. The prediction for the Italian learners was that if they perceive the difference between the two sounds, they should rate words with the incorrect, substituted segment as worse than words with the correct segment. German listeners again served as a control group.

4.4.1 Method

4.4.1.1 Participants

The same participants as in Experiment 2a participated in the explicit rating directly after having finished the eye-tracking task.

4.4.1.2 Materials and Design

In this task, only the critical items and no fillers were presented. To use exactly the same material as in the eye-tracking task, the target words together with up to two words of preceding context were spliced out of the carrier sentences to be presented in isolation, so that listeners could better focus on the words. The preceding context varied depending on the sentence, such as *seinen Handschuh* ("his glove") or *den warmen Anzug* ("the warm suit"; see Appendix III.B.3 in which these contexts are underlined). Note that providing a context is necessary, because the realization of the target segments depends strongly on the context. Consequently, a pronunciation rating of only the target word by itself cannot provide valid results.

During the task, each participant heard each item once with the correct sound and once with the incorrect one, which made a total of 160 trials. For each participant, a different random order was generated and organized as follows: Each participant first heard all words once before they were repeated. In each such block half of the words were presented with the correct sound and half with the replaced segment for each the /h/- and the /ʔ/-items. In the second half, all words were presented in the respective other condition. Words in each half were presented in a different random order for each participant. The rating task took approximately 20 minutes to complete.

4.4.1.3 Procedure

The experiment software was the same as for the eye-tracking experiment and participants performed the task at the same computer, but without the eye-tracking device. Participants were told that they would hear short parts (or phrases) that had been isolated from sentences and see a corresponding picture of the critical word. Their task was to rate how well the depicted word (i.e., the word *Handschuh* "glove" for the phrase *seinen Handschuh* "his glove") was pronounced. For each trial, the picture together with the scale appeared on the screen, and after 1000 milliseconds the sound started to play. Participants had to indicate their rating by pressing one number from 1 to 7 on the keyboard. On the scale, 1 was labelled with *sehr schlecht* ("very poorly"), 3 was labelled with *schlecht* ("poorly"), 5 *gut*

("well") and 7 *sehr gut* ("very well"). After the key-press and an additional of 200 milliseconds, the next trial started automatically.

4.4.1.4 Analysis

Results were again analyzed with linear mixed-effects models using the lme4 package (Bates et al. 2015) in R (Version 3.3.2, R Core Team, 2017). As for the eye-tracking task, words that were unknown to participants were excluded. 226 of 3200 items for the Italians (7.1 %) were removed from the analysis[17]. The dependent variable was the rating which was coded numerically as a number from 1 to 7, where 1 indicated that the listener evaluated the presented word as very poorly pronounced and 7 as very well pronounced, with 2 to 6 as intermediate steps. One overall model was fitted with the fixed factors L1 (coded as German = 0.5, Italian = -0.5), Target (/h/ = 0.5, /ʔ/ = -0.5), Condition (correct segment = 0.5, replaced segment = -0.5) and all interactions. The random-effect structure included intercepts for participant and word (i.e., item) with random slopes for Condition and Target Sound over participant, and Condition and L1 over item which amounts to the full random effects structure (Barr et al. 2013).

4.4.2 Results

Results are given in Table 4.2. Figure 4.4 shows the ratings given by German listeners (left) and Italian learners (right) for the two target sounds and different conditions. The white boxes illustrate ratings of words with the correct sounds and the grey boxes ratings of words with the substituted sounds. As can be seen from the figure, Italian learners rated the words with a replaced segment as much better than the German listeners did, but still worse than they rated the words with the correct sound.

Statistical analyses of the overall model with all factors and both listener groups revealed a significant effect of Condition, with better ratings when the word was presented with the correct segment. Moreover, there was an effect of L1, confirming that Italian learners rated the words overall significantly better than German native listeners. These two factors were involved in three significant interactions. First, the interaction between Condition and L1 confirms that the difference between ratings for the correct vs. replaced condition was larger in the Germans than in the Italian learners. Second, the interaction between Condition and Target indicates that the difference between ratings for the correct

[17] Note that in the explicit task, participants heard each word in both conditions, but only in one condition in the eye-tracking task. Therefore, the number of removed *words* was the same in both tasks (113), but twice as many *items* were removed in the rating task.

vs. replaced condition differed between the two target sounds. The third significant interaction between L1 and Target Sound suggests that /h/-items were rated differently from /ʔ/-items, but this depended on the listeners' first language. However, the three-way interaction between all three factors was not significant.

A follow-up analysis of the German listeners with the same fixed factors as specified above revealed a significant effect of Condition ($b_{Condition}$=4.64, SE=0.18, t=25.56, p<.001; $b_{Intercept}$=4.25, SE=0.08, t=53.86, p<.001), as in the overall model, and an effect of Target (b_{Target}=0.10, SE=0.05, t=2.12, p<.05), with slightly better ratings for /h/-initial items. The interaction between these two was not significant ($b_{Condition:Target}$=-0.13, SE=0.07, t=-1.68, p=.10). An analysis of ratings given by Italian listeners revealed a significant effect of Condition ($b_{Condition}$=0.93, SE=0.35, t=2.68, p<.05; $b_{Intercept}$=5.45, SE=0.18, t=30.81, p<.001), confirming that they rated words with correct sounds as better than words with incorrect, replaced segments. The effect of Target and the interaction between Condition and Target just failed to reach significance (b_{Target} =-0.17, SE=0.13, t=-1.35, p=.18; $b_{Target:Condition}$=-0.18, SE=0.09, t=-2.00, p=.05).

In order to compare these results from the explicit perception task with how well Italian learners produced these sounds, an additional analysis was conducted on the results of the Italian learners' explicit ratings. Note, however, that different sets of learners participated in the production and the perception task, and only a group-wise comparison can be performed. In order to compare results across the different tasks, the rating from the explicit perception task for each substituted form was subtracted from the rating of the word in its correct form, resulting in one value for each participant and each word. In this analysis, we can simply count how often learners perceived the correct pronunciation of a word as better than the wrong one. Note that during the task, only one word was presented at a time, and participants had to rate only the pronunciation of the current version. Their task was *not* to judge which of two words was better or worse. If we henceforth report that one version of a word was rated as better than the other, this is what we compare in the additional analysis.

Overall, words in the substituted condition were rated worse than words in the correct condition in 45% of the cases, they were rated as better in 17% of the cases, and both were rated as equally well in 38%. Looking at the two sounds separately, /h/-items in the substituted condition (i.e., presented with a glottal stop) were rated as better than in the correct version in 19%. This compares well to the production task, in which Italian learners substituted the glottal fricative with a glottal stop in 17% of the cases. /ʔ/-items were rated

as better in the substituted than in the correct condition in 15% of the cases. In the production task, /ʔ/-initial words were produced with a glottal fricative in 11%.

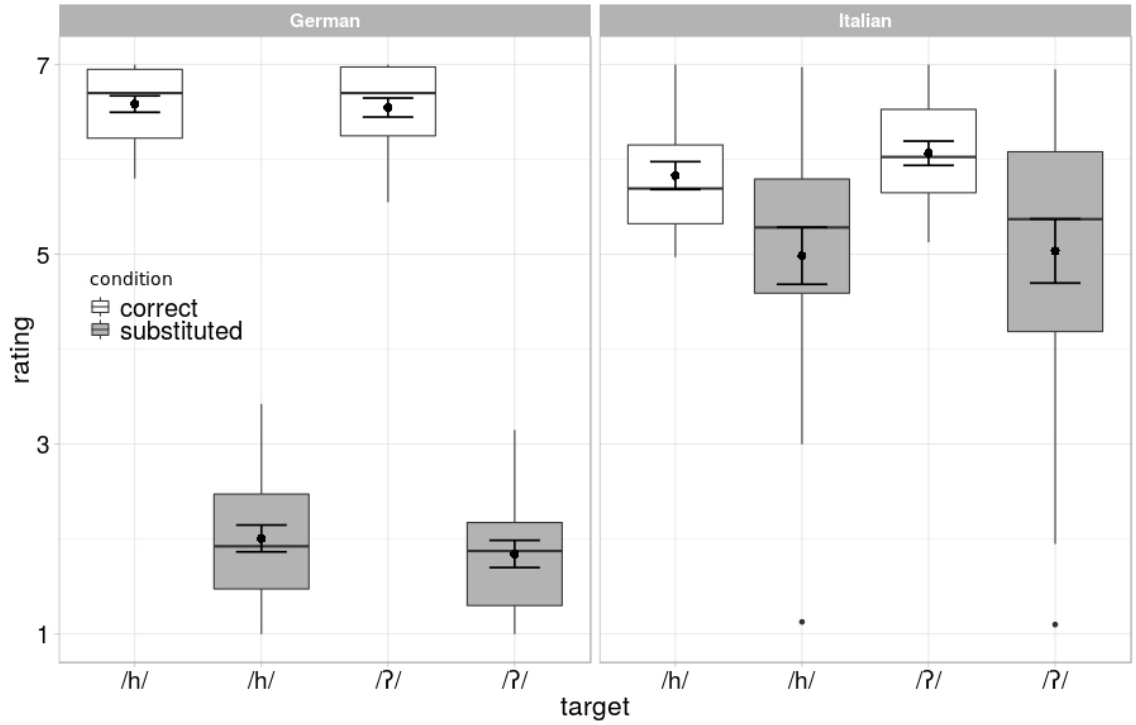

Figure 4.4: Listeners' ratings of /h/- and /ʔ/-items on a scale from 1 (very poorly) to 7 (very well) in the correct (white boxes) and substituted (grey boxes) conditions for native listeners of German (left) and Italian learners (right). Data points are aggregated over words. Dots and whiskers in the boxes indicate means and standard errors.

Table 4.2: Results of the mixed-effects model with listeners' ratings as dependent variable, fitted with Condition (correct, substituted), Target (/h/, /ʔ/), L1 (German natives, Italian learners), and their interactions.

Fixed effect	*b*	SE	*t*	*p*
Intercept	4.85	0.10	49.88	<.001
Target	-0.04	0.08	-0.49	=.62
Condition	2.79	0.19	14.36	<.001
L1	-1.20	0.19	-6.42	<.001
Target:Condition	-0.15	0.06	-2.52	<.05
Target:L1	0.28	0.11	2.41	<.05
Condition:L1	3.71	0.39	9.58	<.001
Target:Condition:L1	0.06	0.11	0.52	=.61

4.4.3 Discussion

The explicit rating task tested whether the finding that Italians were not hindered in spoken-word recognition when presented with the incorrect segment was because they do not perceive the difference between the two sounds. The German listeners again served as a control group. As expected, they rated words with the substituted segment as worse than with the correct segment, and this was true for both targets. Interestingly, there was a small difference between the targets with worse ratings for /ʔ/-items. Even though the interaction was not significant, this difference was presumably due to worse ratings for /ʔ/- than /h/-items in the substituted condition. Not only learners, but also native speakers are typically not aware of the occurrence of a glottal stop in words that are orthographically vowel-initial. Assuming that German listeners deem an orthographic vowel-initial word as starting with a vowel and not /ʔ/, as implied by orthography, this finding may suggest that listeners penalized a supposed "insertion" of /h/ as worse than a /h/-deletion. However, the difference is only subtle and is much smaller than the effect of Condition.

Italian learners judged words that were pronounced with an incorrect, substituted /h/ or /ʔ/ as worse than when pronounced with the correct segment, even though this difference was smaller compared to the German listeners' ratings (i.e., the interaction between Condition and L1). Despite the overall effect of condition in the learners, an additional analysis revealed that only in 45% of the cases, learners rated the correct version of a word as better than the substituted version. This indicates that they did not consistently differentiate between the two sounds (see also the overlap in Figure 4.4 in the Italian learners, but not the Germans). However, note that this task, even if more explicit, is still more difficult than a simple AXB discrimination test, in which listeners focus only on acoustic details and can directly compare the two stimuli in sensory memory. In the current task, there were always intervening stimuli between the correct and the substituted version of a given target. Moreover, since no written information was provided but the identity of the word was only indicated by the picture, listeners additionally had to know which word starts with which sound. This made the task more difficult, compared to a purely acoustic test, which also can be seen in the number of equal ratings for words in both conditions.

Overall, Experiment 2 shows that Italian learners can hear the difference between /h/ and /ʔ/ but are not able to use them for word recognition. This raises the question how Italian learners represent these words in their mental lexicon. The current results leave open

two options. Either Italian learners' representations of German /h/- and /ʔ/-initial words do simply not represent these sounds, or the representations indicate that these words start with some glottal sound, but do not specify whether this is a stop or a fricative. Experiment 3 investigates this by presenting the words with the initial sound deleted instead of exchanged. The perception of /h/- and /ʔ/-initial words with the sounds present or absent was tested in an implicit task in Experiment 3a and in an explicit task in Experiment 3b. If Italian learners have established a common category for these sounds, they should be hindered by a deletion of an initial /h/ or /ʔ/. If, however, the words are represented as truly vowel-initial, the deletion of such sounds, just as their substitution, should not matter for recognition. Again, if being aware of an L2 category matters for its acquisition, Italians should be hindered more when the initial /h/ is missing than when the /ʔ/ is missing.

4.5 EXPERIMENT 3a

4.5.1 Method

4.5.1.1 Participants

For this experiment, 20 Italian learners of German (five males) participated for pay at the University of Munich, Germany. All of the learners spoke Italian as their only first language and none had participated in the first perception experiment. Four of them had participated in the production experiment approximately 8 months earlier. These participants were told that the present task was a different one to what they had done before. The perception experiment was conducted in a different building of the university. The Italian learners were aged between 20 and 37 (mean=28.6, sd=5.2). They started learning German at a mean age of 19.4 (sd=7.6), ranging from 6 to 35 years. Their mean age at the time when they arrived in Germany was 26.0 years (sd=4.3), with the youngest at an age of 18. Even if some of the participants started learning German already at school, before their arrival in Germany no one had longer-term contact to German spoken by native speakers. Whereas some of them were in Munich as exchange students (n=9), others had their permanent residence in Munich. Like the participants in the previous experiments they filled in the questionnaire on their use and self-estimated proficiency in German. A summary of the self-estimated skills in German and the self-estimated accent are presented in Table III.A in the Appendix.

4.5.1.2 Materials

The words and sentences were exactly the same as in Experiment 2 (see Table III.B.3 in the Appendix). New recordings were made with a female native speaker of German. This time the critical segments were not replaced, but omitted, to compare processing of conditions in which the segment is present vs. absent. A different speaker was chosen because the speaker of the items in Experiment 2 had problems in consciously deleting the glottal stop, and, instead of deleting /h/ often substituted a glottal stop. As before, the filler sentences were recorded at least once, and the critical sentences at least three times. Of these, one production with the critical segment and one production with the segment deleted were spliced into a third production which served as carrier sentence. This way, the two resulting sentences were identical and differed only in whether the critical segment, /h/ or /ʔ/, was present or absent. The recordings were cut at negative zero-crossings and after splicing, the duration of the preceding nasal was adjusted. This was done because in some cases in which the speaker omitted the initial sound for the absent condition, she produced longer nasals than when the sound was present. The authors made sure that, despite cross-splicing and adjusting the duration of the nasal, the sentences sounded natural and no audible artifacts remained. For the sentences in which the segments were present, the target onset was defined at the end of the nasal and the beginning of frication for /h/ and at the start of creaky voice for /ʔ/ (see also Experiment 2a *Materials*). For sentences in the deleted condition, the target onset was marked at the end of the nasal.

4.5.1.3 Design and Procedure

The design and procedure of the task were exactly the same as in Experiment 2a: Listeners heard 40 /h/-items, 40 /ʔ/-items, and 40 filler items, all embedded in the carrier sentences described above (see Table III.B.3 in the Appendix). The only difference with Experiment 2a was the manipulation of the phonological realization of the targets, which made use of deletion instead of substitution. Of the critical items, half were presented with the segment present and the other half with the segment absent, but each participant heard each word in only one version.

4.5.1.4 Analysis

Analyses were run only on critical items. Words that were unknown to participants as indicated in a questionnaire were excluded. One-hundred and thirty-three trials (8.3% of the data) was excluded for this reason. Next, trials in which participants did not click on

the picture that corresponded the target were excluded reducing the dataset for another 4.6%.

For the remaining data, target fixations were analyzed, that is, the logOdds-transformed target fixation on the quadrant containing the target picture over the specified time window. The time window was defined between 300 and 800 milliseconds after target onset, as in Experiment 2. Target fixations were analyzed as the dependent variable in a linear mixed-effects model with Target Sound (contrast-coded as /h/=0.5, /ʔ/=-0.5), Condition (correct segment present coded as 0.5, segment absent coded as -0.5), as fixed factors and all interactions. The random-effects structures included intercepts for participant and word (i.e., item) with random slopes for Condition and Target Sound over participant, and Condition over item which amounted to the full random effects structure (Barr et al. 2013), save for the exclusion of correlation parameters between random effects.

4.5.2 Results

The proportions of target fixations are shown in Figure 4.5 for both target sounds; /h/ in the left panel and /ʔ/ in the right panel, and for the conditions in which the words were presented with the segment present (light grey) and in which the segments were omitted (dark grey). Analyses on the critical time-window between 300 and 800 milliseconds revealed no significant effect of Target Sound (b_{Target}=-0.57, SE=0.57, t=-1.01, p=.32; $b_{\text{Intercept}}$=3.48, SE=0.34, t=10.36, p<.001). Moreover, there was no interaction with Condition ($b_{\text{Target:Condition}}$=0.87, SE=0.82, t=1.07, p=.30), but a significant main effect of Condition emerged ($b_{\text{Condition}}$=0.94, SE=0.37, t=2.51, p<.05). Comparing the fixation curves for the different conditions, it seems that this effect is especially due to the /h/-items, as target fixations start to increase later when /h/ was omitted than when it was present. The two lines illustrating target fixations for /ʔ/-items, however, overlap mostly, and appear to diverge only towards the end of the critical time window. This suggests that Italian learners fixated less on the target when the word they heard was produced without the critical segment, and this effect was especially seen in the /h/-items, that is, when the glottal fricative was deleted. The interaction with Target Sound did not reach significance, suggesting that the effect is rather variable over participants. Looking at the fixation proportions it seems that Italians are hindered in spoken word recognition when the initial glottal stop is deleted, but maybe to a lesser extent compared to a deleted /h/.

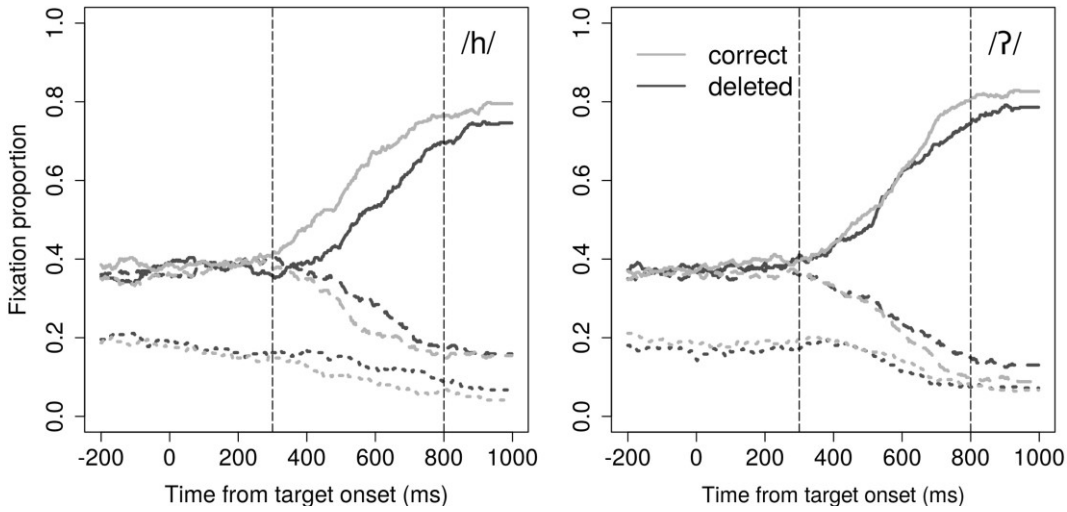

Figure 4.5: Fixation proportions on the targets (solid lines), competitors (dashed lines), and distractors (dotted lines) over time (in ms; 0 is target onset) for /h/-items (left) and /ʔ/-items (right) and in the correct condition (light grey lines) and if the critical segments were deleted (dark grey lines). The vertical lines indicate the time window of analysis.

4.5.3 Discussion

Experiment 3a set out to test whether Italian learners may have established a common representation of "a glottal sound", or alternatively whether they have not established any new category for /h/ and /ʔ/. This question was motivated by the finding that even though Italians heard the difference between the two sounds (Experiment 2b), they did not process them differently in spoken word recognition, as shown in Experiment 2a. Our prediction was that if learners have established any representation, they should look less at the target when the word is presented with the segment missing. We found that Italian learners are indeed sensitive to the deletion of the critical sounds, as they looked less at the target when the word was presented with the critical sound deleted. This effect was numerically larger for the /h/-initial words, but not significantly so. These findings together with Experiment 2a suggest that learners' representations of /h/- and /ʔ/-initial words are specified with a common category of an initial "glottal sound".

4.6 EXPERIMENT 3b

In order to test whether Italian learners would perceive the difference between words with the initial segments present vs. absent when asked explicitly, the same type of task as in Experiment 2b was administered, but this time words were presented with the critical

segments present or absent. If learners hear the difference between presence and absence of the critical sound, they should rate words without the segment as worse than with the (correct) segment present. Moreover, if explicit knowledge about a sound category has an impact on establishing a mental representation for an L2 sound, learners should rate words in which /h/ is missing as worse than their correct form, and this difference should be larger than when the /ʔ/ is missing.

4.6.1 Method

4.6.1.1 Participants

The same participants as in Experiment 3a participated in the explicit rating directly after having finished the eye-tracking task.

4.6.1.2 Materials and Design

Targets with a minimal preceding context were spliced out of the sentences that were presented in the eye-tracking task. Each participant was presented each item, once with the critical segment present and once with the segment deleted, which made a whole of 160 trials. The randomization was applied as in Experiment 2b, so that each word was presented with either the sound present or absent in the first half, and a second time in the second half of the experiment, but this time in the respective other condition.

4.6.1.3 Procedure

The design and procedure of the task were exactly the same as in Experiment 2b: Participants had to indicate their rating by pressing one number from 1 to 7 on the keyboard. Again, 1 was labelled with *sehr schlecht* (“very poorly”), 3 with *schlecht* (“poorly”), 5 with *gut* (“well”) and 7 with *sehr gut* (“very well”). After key-press and an additional of 200 milliseconds, the next trial started.

4.6.1.4 Analysis

Words that were unknown to participants were removed, which were 264 items (8.3% of the data). Ratings were analyzed using linear mixed-effects models with the full random effects structure. The dependent variable was the rating given for the word by the participants, which was coded numerically as a number from 1 to 7, where 1 indicates that the listener evaluated the presented word as very poorly pronounced and 7 as very well pronounced, with 2 to 6 as intermediate steps. One model was fitted with the fixed factors Target Sound (/h/ = 0.5, /ʔ/ = -0.5) and Condition (correct segment present = 0.5, segment

absent = -0.5), and an interaction between these. The random-effect structures included intercepts for participant and word (i.e., item) with random slopes for Condition and Target Sound over participant, and Condition over item (Barr et al. 2013; i.e., within participant: Condition and Target Sound, within item: Condition).

4.6.2 Results

Figure 4.6 illustrates the ratings given for words with the two target sounds, and in the different conditions (white boxes for the (correct) sound present, grey boxes with the critical sound absent). Analyses revealed a significant effect of Condition ($b_{Condition}$=0.76, SE=0.14, t=-5.41, p<.001; $b_{Intercept}$=5.66, SE=0.14, t=41.43, p<.001), confirming that Italian learners of German rated words with the initial sounds present as better than when the segments were omitted. Moreover, the effect of Target (b_{Target}=-0.76, SE=0.13, t=-5.85, p<.001) and the interaction between Condition and Target were significant ($b_{Target:Condition}$=0.96, SE=0.26, t=3.74, p=.01). The effect of Target Sound indicates that overall, /h/-items were rated as worse. The interaction indicates that the difference between ratings with and without initial segment was larger in /h/- than /ʔ/- items, as can also be seen in Figure 4.6. Two follow-up analyses with the same fixed factors as specified above for the two target sounds separately revealed that, for /h/-items, there was a significant effect of Condition ($b_{Condition}$=1.23, SE=0.26, t=4.75, p<.001; $b_{Intercept}$=5.28, SE=0.17, t=31.09, p<.001). The effect of Condition reached also significance in the analysis of the /ʔ/-items ($b_{Condition}$=0.28, SE=0.08, t=3.70, p<.01; $b_{Intercept}$=6.05, SE=0.13, t=46.29, p<.001), but, given the interaction, is significantly smaller than for /h/.

In order to calculate how often learners rated a given word in the correct condition as better than when it was presented with the critical sound deleted, for each participant and each word, the rating of the word in the deleted condition was subtracted from the rating given for the word when presented in the correct version. Words with the correct segment present were rated better than the words with the critical segment deleted in 44% of the cases, they were rated as worse than words with the segment deleted in 14%, and both were rated equally well in 42%. Analyzing the two target segments separately revealed that /h/-items in the deleted condition were rated as better than the word with the target sound present in 11%, which compares to the production task in which Italians deleted the /h/ about 9%. Of the /ʔ/-items, 16% were rated as better in the deleted than in the correct condition. In the production task, Italian learners deleted the glottal stop in 24%.

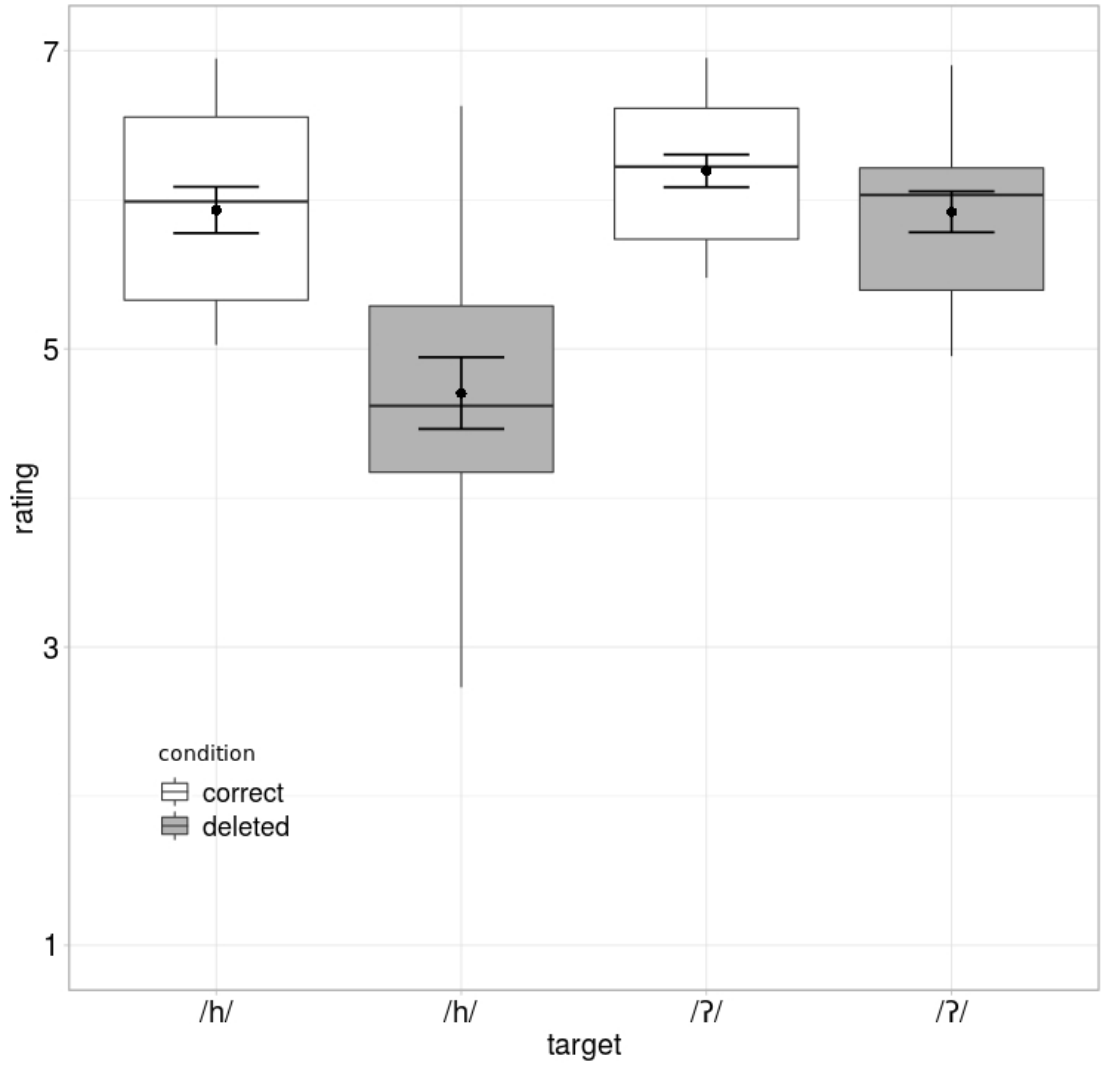

Figure 4.6: Listeners' ratings of /h/- and /ʔ/-items on a scale from 1 (very poorly) to 7 (very well) in the correct (white boxes) and deleted (grey boxes) conditions for the Italian learners. Data points are aggregated over words. Dots and whiskers in the boxes indicate means and standard errors.

4.6.3 Discussion

The explicit rating task tested whether Italian learners would hear the difference between words with the target segments present compared to when they were deleted. Results showed that learners indeed perceived the difference, as evidenced by worse ratings for words in which the critical segment was deleted. Even though there was still overlap of the ratings (equal ratings for both versions in 42%), the effect of condition could be observed for both segments. Moreover, the effect was stronger when the glottal fricative in /h/-items

was absent, which is in line with the findings in Mitterer and Reinisch (2015) that orthography affects explicit judgments.

The knowledge that /h/- and /ʔ/-initial words do not simply start with a vowel is hence stronger for /h/. This may be due to the fact that for words like *Helm* ("helmet") both orthography and formal training indicate that it is not a vowel-initial word. This is not the case for words like *Apfel* ("apple").

4.7 General Discussion

The present study was based on two pillars. First of all, the two glottal sounds in German, /h/ and /ʔ/, form a sound contrast that may be difficult to acquire for learners with a Romance L1 background. Second, there is a dearth of data on L2 sound learning with two new L2 sounds that are unlike any L1 sound categories. More specifically, we asked how well German /h/ and /ʔ/ are acquired by Italian learners in perception versus production, whether there is an advantage to acquire /h/ due to its orthographic coding and explicit treatment in education, and to what extent there are differences in explicit versus implicit perception tasks, also with regard to any asymmetry in acquisition.

Experiment 1 focused on production and showed that the sound pair is problematic, and it is especially so as a sound contrast, that is, the two sounds are confusable. It was found that although Italian learners performed relatively well overall, they were far from being near-native, with 70% target-like productions. Although /h/ was overall produced correctly more often than /ʔ/, the majority of "vowel-initial" words were still appropriately produced with a glottal stop. An analysis of Italian speakers' productions in their own L1 further showed that this is due to acquisition and not transfer from the L1. In their L1, the Italian learners produced vowel-initial words overwhelmingly without glottalization. Regarding the two sounds as a contrast, Italian listeners produced a sizeable number of substitutions of these sounds - in both directions - indicating that they acquire them as similar.

Experiment 2 and 3 focused on perception and demonstrated that Italian learners of German are hindered in spoken word recognition when the initial /h/ or /ʔ/ of a German word was deleted (Experiment 3a). That is, the presence of the segments helped learners recognize the words, although neither segment exists as sound category in Italian. However, which of the two sounds was present did not make any difference for the learners (Experiment 2a), even though they were able to differentiate them in an explicit task (Experiment 2b). This discrepancy is in line with previous findings that learners perform

better when asked explicitly to differentiate sounds compared to implicit tasks that test processing in more natural situations (e.g., Díaz et al. 2012; White et al. 2017). Moreover, we showed that being aware of the unfamiliar sound helps the learner establish a new L2 category, since /h/ overall seems to be somewhat more robustly acquired than /ʔ/. This advantage for /h/ was evident in the rating results of Experiment 3b, in which deleting /h/ led to a stronger reduction in pronunciation ratings than the deletion of /ʔ/. A similar numeric effect in the eye-tracking data of Experiment 3a, however, was not significant, showing that the advantage for /h/ is less pronounced in an implicit word-recognition task than in an explicit rating task.

With regard to the relation of perception and production, the results are somewhat surprising, given that the relatively high number of 70% correct realizations stands in contrast to the finding that sound substitutions did not hinder learners in spoken word recognition in the implicit task in Experiment 2a. This raises the question whether the acquisition of /h/ and /ʔ/ is a case in which production precedes perception (e.g., Flege & Eefting, 1987; Sheldon & Strange, 1982).

Despite the similarity in terms of overall L2 abilities of the participants in the first two experiments, the results indicate that production is in fact closer to the L1 target than perception. There are at least two different explanations for this. First, the discrepancy between the production task and the implicit perception experiment with sound substitutions may be attributed to the different types of tasks that put different demands on the learner (Díaz et al. 2012). The eye-tracking task in Experiment 2a tested implicit online processing of spoken words. The production task may trigger a more conscious speaking and listening mode, in a similar way as the explicit rating task. Even though participants were not presented with written words in the production task, they may have accessed orthographic knowledge of the learned words and put more effort into producing the correct sounds. This finding illustrates once again the effect of task type which may trigger different paths to access representations (Krieger-Redwood et al. 2013). One could expect that in spontaneous speech, when learners are not in a laboratory situation, they may perform slightly worse than in the production task in the present study.

This explanation does, however, not straightforwardly explain why participants perform better in the production task than in the explicit perception task. One could assume that in terms of explicitness, the production task in the present study lies somewhere between the two perception tasks; yet participants produced the correct pronunciation in about 70% of the cases but preferred the correct pronunciation in the rating task in only

about 40% of the cases. An alternative possibility then is that, when learners can hear the difference between two new L2 sounds, they can apply this knowledge more easily in production than perception. That is, the better performance in the production task reflects a truly better ability to acquire discriminable speech sounds in production than in perception (see also Flege & Eefting, 1987; Sheldon & Strange, 1982) and cannot be explained by task differences.

However, it has to be noted that this concerns a pair of sounds that are used in para-linguistic functions in Italian, so that learning new motor routines is not really necessary. As reviewed in the Introduction, the glottal fricative is used in laughing or sighing in Italian. Glottalization can be produced in initial position in vowel-initial words, but only in hyper-articulated speech. In addition, glottalization can be used to signal lexical stress. When speaking and hearing in an L2, learners often transfer patterns from their L1 to the L2, thereby using cues that may differ from how native speakers would typically speak (Bohn, 1995; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; McAllister, Flege, & Piske, 2002; Yamada, 1995). However, sourcing cues from one's L1 can also be beneficial. For instance, learners have been shown to transfer their use of cues of a familiar contrast to an unfamiliar position in the word (Broersma, 2005) or to reassign a familiar cue to a different function in the L2 (Eger & Bohn, 2015). Parallel to that, the use of glottalization in Italian, albeit differently from its use in German, may nevertheless foster its acquisition, and a similar type of transfer may be found for /h/. In contrast, we would expect that articulatorily difficult sounds (such as apical trills) may be more easily incorporated in perception than in production. Nevertheless, the current data indicate that the ability to distinguish two new L2 speech sounds whose production routines are already known from non-linguistic domains may be integrated more efficiently in speech production than in perception. This would also indicate that L2 perception and production, once sounds are auditorily distinguished, may progress quite independently.

The second question was whether /h/ would be more robustly acquired than /ʔ/, due to its orthographic coding and well-documented influences of orthography on L2 acquisition (Bassetti, 2017; Escudero et al. 2008), as well as the fact that /h/ tends to be covered during teaching while glottal stop is not. Some asymmetric patterns - even though not always statistically significant - were found in the production task in Experiment 1 and also in Experiment 3, in which we tested deletion costs of word-initial /h/ and /ʔ/. The fact that some differences were found supports the available evidence that L2 learning is open to orthographic influences. However, given the expected massive a-priori advantage for /h/

over /ʔ/, these differences were surprisingly small. The glottal stop was certainly not neglected, neither in production nor in perception, even though there is no prevalent use of glottalization before word-initial vowels in Italian - as also evidenced in the present production task. Considering that speakers typically do not even have a notion of this sound, this finding is noteworthy. This is even more so, since the literature indicates that the sensitive period for language learning may end earliest for phonetic and phonological aspects of language learning (Pallier, Bosch, & Sebastian-Gallés, 1997). Our data show that learners are nevertheless still able to acquire a new sound without explicit instruction. Part of this success may be attributed to the fact that glottalization is used in Italian, but in a different function, as argued above. Additionally, the learners in the present study had every-day contact to German spoken by native speakers, and were hence passively exposed to productions of the glottal stop in "vowel-initial" words.

However, since glottalization and breathy sound quality (in /h/) are not contrastive in Italian, the glottal stop is not fully differentiated from /h/ in the learners' lexicon. Therefore, learners were not hindered in spoken-word recognition when the sounds were substituted, they did not differentiate them in all cases in the explicit tasks, and they confused them in production, which never happened for the German control group. This remaining overlap between the categories may further be enhanced by hearing inaccurate productions of Italian learners of German, including their own (Eger & Reinisch, 2019a).

Our third question was whether, in perception, there is a difference in performance between explicit and implicit tasks. This was certainly the case. For word recognition, learners did not make use of the /h/-/ʔ/ distinction (Experiment 2a), even though they were able to make use of the difference in the explicit task (Experiment 2b). As such, the current data reinforce that L2 perception of new sounds may dissociate between the phonetic and lexical level (Díaz et al. 2012), so that auditory discrimination does not automatically translate to use for lexical processing. Moreover, in Experiment 3, the explicit task revealed a significant preference for /h/, while the implicit task did not. This aligns with data from German native speakers in that explicit tasks are more likely to elicit orthographic effects (Mitterer & Reinisch, 2015).

However, the data also differ from the findings in Mitterer and Reinisch (2015) by revealing at least a numeric tendency for a preference for /h/ in the eye-tracking data of Experiment 3a. It is important to consider that the population investigated differed between the two studies. In the previous study, only native speakers were tested, and it can be expected that they already had well-defined, separate mental representations of the critical

sounds. Whether an L1 sound is indicated by a letter or not, as in the case of /h/ vs. /ʔ/, did not make a difference for online processing. Italian learners, by contrast, presumably did not have a well-defined representation of these two sounds when they started learning German, as none of these segments is used as a speech sound in Italian. The different status of these two sounds, including the orthographic coding and thus awareness of /h/ but not /ʔ/, may shape the formation of a representation of new categories when learning German. The resulting representation may then mediate the effect on processing for L2 learners, rather than being due to an online influence of orthography in processing.

Overall, these data indicate how a sound contrast that is completely unlike the categories of the L1 can be learnt. Our data concur with the classic finding that such sound contrasts can be distinguished at an auditory level (Best et al. 1988), given that listeners were sensitive to the difference in the explicit rating task. Our data additionally indicate that this auditory sensitivity may be utilized to guide production before it is used to guide spoken-word recognition.

It is worthwhile to consider the implications of these results for the phonological status of glottal stop in German. Our data indicate that Italian learners of German tend to confuse /h/ and /ʔ/ in German. This contrasts with the classical phonological treatments, which view /h/ and glottal stop as completely different entities, one as a phoneme of the language, and the other as an epenthetic segment (Wiese, 1996). However, this categorical difference at the phonological level is clearly not apparent in the language, otherwise, Italian learners should have acquired the sounds rather differently and not find them confusing. This, in turn, questions the rationale to treat /h/ and glottal stop as two completely different entities in the phonology of German.

Summing up the present study, we found that Italian learners of German can perceptually differentiate two sounds that do not exist in their L1 when asked explicitly, but at the same time they have difficulties in differentiating them in online-processing of speech. When the segments are absent, listeners are hindered in spoken word recognition, but this effect depended on the status of the segment, that is, whether listeners are aware of the existence of this sound. In production, learners realized both sounds correctly in most cases, but did not reach a native-like level. Regarding our question of how two unfamiliar sounds – that have no clear counterpart in the L1 – may be acquired and represented, the results suggest that learners acquire them differently in perception and production. In perception, they establish one new category onto which they map both sounds, with /h/ being more dominant. In production, they are better able to make use of the ability to

discriminate the sounds on an auditory level and are closer to the native level than in perception.

5 SUMMARY AND CONCLUSION

The overall goal of this thesis was to investigate some factors that may be responsible for the difficulty in overcoming an accent in a foreign language. Nonnative speech naturally differs from that of native speakers. This can mainly be attributed to speech patterns of the learner's L1 that influence production and perception of second language sounds. Previous research has demonstrated that different types of L2 sounds are differently easy to acquire (Best & Tyler, 2007; Flege, 2003; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008). In addition, several factors have been identified that affect the strength of an accent (Gluszek, Newheiser, & Dovidio, 2011; Moyer, 2007). However, less attention has been paid to the question of why this accent appears to be so persistent. The experiments described in this thesis aimed at understanding some facets of this difficulty.

Foreign accent in learners' speech productions is reliably detected by native speakers of that target language (Abrahamsson & Hyltenstam, 2009; Flege, 1984). This is also because properties typical of foreign accent can be found on various dimensions of spoken language (Anderson-Hsieh, Johnson, & Koehler, 1992). Segmental deviations in words, such as sound deletions or substitutions, have been shown to contribute to a perceived foreign accent, even though supra-segmental properties over the whole utterance also play an important role (Anderson-Hsieh et al. 1992). The present thesis focused on the segmental level in the production and perception of second language sounds. This offered the possibility to gain insights into how single L2 *words* with unfamiliar sounds may be represented in the mental lexicon. These representations, in turn, are likely to be the phonolexical units that are used to recognize spoken words and judge the goodness of pronunciation, and also likely as a basis for speech production.

Chapter 2: Acoustic Cues and Proficiency in Accent Perception

The experiment in Chapter 2 asked whether learners perceive L2 words spoken with a foreign accent that matches their own L1 as good productions. Previous research has demonstrated that nonnative listeners can have an advantage over native listeners in understanding foreign-accented speech when speaker and listener share their L1 (Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, Bradlow, 2008). Moreover, spoken-word recognition is facilitated by hearing words produced in the accent typical of one's L1 (Weber, Broersma, & Aoyagi, 2011). However, a question that has received much less

attention is whether accented words also sound acceptable to L2 learners, and whether this is modulated by the learners' proficiency in and experience with the L2.

To address this question the experiment reported in Chapter 2 investigated listeners' sensitivity to foreign accent in L2 words in which the accent matched their own L1. German learners of English and native English speakers used as a control group were asked to rate the goodness of English words that had been recorded by another group of German learners. Words belonged to minimal pairs differing in one of three sound contrasts that are typically minimized by German learners. The strength of the accent – or reversely, the goodness of the pronunciation – differed between words with easy or difficult sounds (easy vs. difficult words). "Easy" words contained those sounds that are familiar in the L1, whereas "difficult" words contained sounds that are unfamiliar in German as a category or in a given position of the word. The latter were thus expected to be produced less accurately and sound more accented. In fact, differences between the productions of easy vs. difficult sounds were found by means of acoustic analyses. The accent strength also differed according to the overall production quality of the material sets (good, intermediate, poor), which was determined as the produced difference between easy and difficult words within minimal pairs. The listeners' task was to rate on a seven-point scale how well the presented word was pronounced.

Results showed that the native listener control group judged the pronunciation of the difficult words as worse than the easy words, as expected. This difference was larger in the intermediate and poor material sets, whereas easy and difficult words were overall judged as similarly good in the best material set. Looking at the learners, results revealed that the goodness judgments depended on a combination of material and the learners' own proficiency in their L2. The more proficient the learners were and the more experience they had with spoken English, the more they perceived the difference between the pronunciation of easy and difficult words, reflected in worse ratings for the latter. This pattern was clearer, as for the native listeners, in the poorer material sets, where the cues to the contrasts were smaller and therefore easy and difficult words sounded more similar to each other. The less experience learners had and the less proficient they were, by contrast, the less they perceived the difference between easy and difficult words. This pattern did not change across the material sets as much as for the higher-proficiency learners. In other words, the lower-proficiency learners were more likely to accept both less and more strongly accented words as equally good instances of English words. This was true even for words that were

pronounced with a stronger accent. The higher-proficiency learners, by contrast, were more sensitive to the accent, patterning more with the native listeners.

If learners judge a word spoken with an accent as good, this indicates that they perceive a reasonably good match with a representation of this word. In this sense, the results suggest that learners' representations of L2 words with difficult sounds are shaped by the accent typical of their own first language, but these representations can become "less accented" and more native-like with contact to native speech.

Chapter 3: A Self-Benefit for Spoken-Word Recognition in L2

The experiments reported in Chapter 3 extended the investigation of whether familiarity with specific speech patterns in a second language shapes the learners' mental representations. If familiarity with foreign-accented speech has an influence on how listeners perceive L2 words, learners may be also better at recognizing words produced by themselves. This is because they may be even more familiar with their own productions in a second language than with production patterns of other learners. A self-benefit for recognizing L2 words may work in principle as the interlanguage intelligibility benefit (e.g., Bent & Bradlow, 2003), but it may apply to one's own very specific production patterns. Moreover, previous findings have indicated that the intelligibility benefit is especially found among low-proficiency L2 learners; that is, when low-proficiency L2 listeners are presented with productions of low-proficiency speakers, who typically speak with a stronger foreign accent and produce cues to difficult L2 contrasts less clearly (e.g., Hayes-Harb et al., 2008; Pinet, Iverson, & Huckvale, 2011; van Wijngaarden, Steeneken, & Houtgast, 2002; Xie & Fowler, 2013). Based on that, the questions arose as to whether a self-benefit may be larger for low-proficiency learners, and how it may be related to the availability of acoustic cues.

To test these questions, two experiments were conducted in which learners had to identify words from difficult minimal pairs of the same type as in Chapter 2. In the first experiment, learners' identification accuracy of foreign-accented words was compared between when presented with their own vs. others' productions. Based on how well learners had produced the difficult contrasts, they were assigned to one of three proficiency groups (good, intermediate, poor). In the perception experiment, one's own voice was matched with a set of proficiency-matched unfamiliar speakers. The aim of the second experiment was to test the relation between learners' production and perception skills, and whether this depends on the quality of the material. Therefore, speakers/listeners from the best and

poorest group performed the same type of task as in the first experiment, but this time they were presented with productions from the opposite proficiency group (i.e., best speakers heard productions from the poorest speaker group, and the other way round).

Results showed that a self-benefit could be found: Learners were indeed better at understanding spoken words in a second language when they heard their own productions than productions of other proficiency-matched learners. This benefit did not differ significantly between the different proficiency groups. However, analyses of the sound types (easy vs. difficult words) in combination with the learners' proficiency suggested that the self-benefit was found especially under difficult listening conditions, that is, when only few and small cues to the critical contrasts were available. This indicates that L2 learners adapt to their own, accented production patterns in a second language.

The second experiment showed that better producers in L2 make also better perceivers, but the pattern was complex: Whereas both high- and low-proficiency learners benefit from the availability of clearer acoustic cues in the material, learners with better production skills could perceptually exploit the cues to a larger extent. An additional comparison with the self-benefit found in the first experiment revealed that acoustic cues to the contrasts were overall more important for correct identification than perceiving one's own voice. However, when the cues were sufficient for identification to a certain extent, the self-benefit was on top of that. These findings suggest that listeners benefit from acoustic cues especially when they also realize the contrasts in production, while the self-benefit is found additionally within groups of the same proficiency.

Chapter 4: The Role of explicit Knowledge in the Acquisition of two novel Sounds

The experiments reported in Chapter 4 explored the role of explicit knowledge of a difficult L2 sound, as mediated by orthography, in the acquisition of two unfamiliar sounds that have no obvious counterparts in the learners' L1. In order to test this, it was important to select two sounds of which one is widely known by learners and the other one is typically not, but that are similar in terms of being a difficult L2 sound. To gain a more comprehensive picture of how learners master two L2 sounds that are dissimilar from any L1 category, the experiments reported in Chapter 4 tested Italian learners' acquisition of German /h/ and /ʔ/, of which only /h/ is orthographically coded and typically known by learners. These sounds offered a good possibility, because they share some acoustic and distributional properties and have been shown to be similarly important for spoken-word recognition in native listeners (Mitterer & Reinisch, 2015). Moreover, the sounds are not

used as sound categories in Italian, but can occur in paralinguistic function (Bertinetto & Loporcaro, 2005).

Prior studies that investigated L2 sounds that are dissimilar from any L1 category often looked at production or perception only (e.g., White, Titone, Genesee, & Steinhauer, 2017; Zimmerer & Trouvain, 2015), dealt with two unfamiliar sounds that may be mapped onto different L1 categories (e.g., Scarpace, 2014), or concentrated on acoustic discrimination but not on lexical aspects (e.g., Best, McRoberts, & Sithole, 1988). One of these works has demonstrated that learners can be surprisingly good at discriminating sounds that are not similar to any native category, such as click contrasts in Zulu for native speakers of American English[18] (Best et al. 1988). Yet, listeners perform typically better in more explicit discrimination tasks than in implicit tasks that resemble word recognition in more natural situations (e.g., Llompart & Reinisch, in press_a; Sebastián-Gallés & Díaz, 2012; White et al. 2017). This is mainly because in the former, learners can focus on acoustic differences whereas the latter task requires learners to have encoded these sounds into L2 words, and access them quickly and more spontaneously.

Production as well as perception in an implicit task and in an explicit task were examined. In Experiment 1, semi-spontaneous productions were elicited. To investigate perception, a set of experiments with other learners of a comparable proficiency level tested how learners perceive words with the correct vs. the substituted segment (Experiment 2), or words with the correct segment present vs. absent (Experiment 3). For the implicit task, the eye-tracking method was used in order to test how learners process spoken words including the target sounds. Assuming that listeners evaluate potential lexical candidates for the speech input they are hearing, results from this task can give insights into how words are represented in the mental lexicon (Huettig, Rommers, & Meyer, 2011). For the explicit task, learners had to judge how well the words were pronounced. Throughout the experiments, no written words were presented, but only pictures in order to test the impact of being aware of a sound, and minimize the immediate effect of orthography.

Results of the production task showed that learners realized both sounds correctly in about 70 % of the cases, in spite of the fact that none of them occurs as distinctive sound category in their L1. Among the non-target-like realizations, both sounds were substituted with each other similarly frequently, but the glottal stop was deleted more frequently than

[18] Note that sounds without clear counterpart in the L1 can be further subdivided into un-categorized sounds vs. sounds that are not even perceived as speech, such as clicks. Both types are typically not assimilated to existing L1 sounds. For more details, see Best and Tyler (2007).

the glottal fricative. The implicit perception tasks showed that listeners were hindered in spoken-word recognition when the sounds were deleted, especially when /h/ was missing, but not when the sounds were substituted with one another. The explicit tasks showed that learners could acoustically differentiate between the sounds. Deletions of both sounds were also perceived, but /h/-deletions were associated with lower goodness ratings.

The findings suggest that learners have established one common fuzzy category to which both sounds map in spoken-word recognition, which can be differentiated better when asked explicitly and in production. The differences between the target sounds in the different tasks indicate that /h/ is represented better, suggesting that explicit knowledge (as mediated, for instance, by orthography) can help establish a new sound category.

Conclusions, implications, and future directions

Two main conclusions can be drawn from the findings, which also have implications for L2 learning and teaching. The first conclusion is that learners' representations of L2 words are not only non-target-like due to interferences from the L1 phonetic-phonological system. Rather, representations of L2 words are shaped by the accent typical of the learner's L1, and additionally by the learner's own specific production patterns. In addition to the perceptual difficulties L2 learners have in differentiating novel sounds and contrasts in initial stages of L2 acquisition, they are also frequently exposed to nonnative productions (including words with these difficult sounds) from fellow learners, themselves, and often even teachers.

The notion that frequency of specific pronunciation patterns in the input influences how speech is perceived is in line with findings on variant frequency effects in L1 (e.g., Connine, 2004; Connine, Random, & Patterson, 2008). For instance, American English listeners were more likely to perceive a form as a word in a continuum between a real and a non-word if the carrier stimulus contained a flap instead of a voiceless alveolar stop (Connine, 2004). The flap is the pronunciation variant of /t/ that is predominantly used in American English (Patterson & Connine, 2001). The author proposed that highly frequent variants are stored in the mental lexicon alongside with canonical forms, and may be even dominant and accessed directly (Connine, 2004). If it is true that input frequency influences the mental lexicon, this should also be true for nonnative listeners. Actually, it may play an especially prominent role for this listener group, since in an initial state of acquisition, L2 representations in learners are not target-like anyhow (e.g., Cutler, 2015). Frequent

exposure to accented speech and limited native input likely adds to the "accent" in L2 lexical representations.

This would also account for the matched intelligibility benefit, that is, for the finding that foreign-accented speech can be as intelligible as native speech – or even more intelligible – when talker and listener share their first language background (e.g., Bent & Bradlow, 2003). Additionally to retrieving "phonetic and phonological knowledge" (Bent & Bradlow, 2003, p. 1607), nonnative listeners may perceive a reasonably good match between accented input and their reference representations. However, an important point to be made is that also native listeners of a target language can perceptually tune into foreign-accented speech, even within a few minutes of exposure to it (e.g., Clarke & Garrett, 2004; Bradlow & Bent, 2008; Witteman, Weber, & McQueen, 2013). Yet, whereas – due to their perceptual flexibility – they can become better at understanding accented speech, it is less clear under which conditions they would perceive foreign-accented words as *acceptable* (but see also, e.g., Thompson, 1991). More importantly, being less sensitive to an accent over time due to experience with that accent may not have major consequences for native speakers, but it may be detrimental for developing L2 pronunciation in learners: If the accent is not noticed, learners may not notice a need to change.

The findings further strongly suggest that – above and beyond adapting to general accent characteristics of one group with a certain L1 background – L2 learners additionally adapt to their own, very specific production patterns. This finding is best explained by a model that incorporates context-dependent perception (e.g., Kleinschmidt & Jaeger, 2015). Different representations may be targeted depending on speaker or group membership. When hearing unfamiliar voices, listeners may make use of more general acoustic cues, but access very specific subsets of representations in their own voice once they recognize it[19]. This is in line with previous findings that perception of one's own voice differs from the perception of others' (Graux, Gomot, Roux, Bonnet-Brilhault, Camus, & Bruneau, 2013; Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008; Xu, Homae, Hashimoto, & Hagiwara, 2013). Importantly, the present findings contribute to the existing literature in that they show a self-benefit not only for voice recognition (*who* is speaking), but in relation to phonetic-phonological awareness in an L2 (*what* is being said). Crucially, as outlined

[19] A study on word identification in L1 of own and others' productions revealed that under conditions in which listeners could *not* recognize themselves, they were better at understanding words produced by a speaker whose speech characteristics were close to the average of the speaker community (Schuerman, Meyer, & McQueen, 2015).

throughout the thesis, non-canonical speech patterns due to foreign accent often have consequences on the lexical level. That is, non-target-like production patterns can diminish the acoustic differences of phonemically relevant contrasts and thus make word recognition more difficult.

How could a self-*benefit* be problematic? The answer is, whenever control mechanisms do not work properly. During speaking, the feedback system monitors the (auditory and sensorimotor) input, as outlined for instance in the DIVA model (Tourville & Guenther, 2011). The idea is that the incoming feedback is compared to forward models of the intended output. If there is a mismatch, the current production configurations are adjusted to align these two (e.g., Houde & Jordan, 2002; Niziolek, Nagarajan, & Houde, 2013). When speaking a second language, especially in the beginning, monitoring one's non-target-like productions may not evoke a mismatch. This is because representations are shaped by the general accent of one's L1 speaker community (Chapter 2) and by one's own speech patterns (Chapter 3). In this light, the findings are in line with studies on children with a phonological disorder. These were worse at detecting own erroneous productions compared to other children's utterances, presumably because their representations are imprecise due to extensive exposure to own erroneous productions (Shuster, 1998; Strömbergsson, Wengelin, & House, 2014). This acceptance or lowered awareness of own non-target like productions can be seen as parallel to the L2 learners in the present study: While the children with the phonological disorder accepted their utterances as correct despite the mispronunciations, the L2 learners recognized their own productions despite the accent, better than in others.

Looking at the findings from this perspective, the benefit of better recognition of oneself and a lowered awareness of own "errors" may be just two sides of the same coin: Both are the result of familiarity with own accented speech patterns and the tight link between production and perception in individuals. Again, if there is no mismatch perceived, the need to adjust pronunciation is not obvious. Being repeatedly exposed to these non-target-like productions, adaptation to own production patterns may become even stronger, and this circle may repeat itself. This would contribute to explaining why L2 learners do not overcome their accent even after many years of practice, and also why sensitivity to acoustic properties of own accented speech is lowered compared to others. Note that the experiment in Chapter 3 showed a self-benefit for understanding L2 words, strongly suggesting that learners adapt to own accented, non-target-like production patterns in an L2. That this is directly linked with awareness of one's own accent has yet to be shown.

Even though prior studies addressed self-perception and online control of speech in L2 learners (e.g., Baker & Trofimovich, 2006; Howell & Dworzynski, 2001; Mitsuya, Samson, Ménard, & Munhall, 2013), more research is necessary to better understand how a self-benefit in understanding foreign-accented words, online monitoring of speech, and awareness of one's own accent are linked. The findings of Chapters 2 and 3 offer a valid basis for future research. If it is the case that better recognition of oneself and lowered awareness of one's own accent are two sides of the same coin, repeatedly listening to own productions may not help, rather, it may even harm. Related to that, it has been shown that learners were better at repeating L2 words with specific stress patterns when they heard their own *corrected* productions than a native model (Bissiri & Pfitzinger, 2009; see also Peabody & Seneff, 2006). Even though this seems promising, in classroom situations, laborious methods of this kind are typically not practicable. What is feasible, however, is providing learners with more native input, so that they are not mainly exposed to accented productions, including their own.

The second conclusion that can be drawn from the findings is that explicit knowledge of the existence of an L2 sound can help establish a category for this sound, but it is not absolutely necessary. Chapter 4 revealed that explicit knowledge, as mediated by orthography, can have an effect on L2 perception and production, which corroborates earlier findings (e.g., Bassetti, 2008; Escudero, Hayes-Harb, & Mitterer, 2008). Crucially, since no written words but only pictures were used, the differences found between the target sounds are not due to an immediate influence of reading. Rather, the findings suggest that during the process of acquiring words in an L2, orthography can raise the learners' attention towards certain sounds via sound-to-orthography mapping (van Orden, 1987). This awareness, in turn, can be reflected in L2 words' mental representations. This is in line with studies showing that orthography impacts novel word learning (Escudero et al. 2008), and that the same learning effect can also be achieved by drawing learners' attention towards articulatory information on the critical contrast (Llompart & Reinisch, 2017).

Since /h/ is usually known by learners but they are typically not aware of the existence of /ʔ/ (and none of the sounds is used contrastively in the learners' L1, Italian) one prediction was that learners may have a clear advantage for the glottal fricative. Results indeed point to a dominance of this sound in learners' representations, but this did not manifest itself in the same way in all tasks. In production, the glottal stop was deleted more frequently than the glottal fricative, but there was no overall advantage for /h/, and substitutions occurred similarly frequently in both directions, indicating that the sounds are

acquired as similar (Experiment 1 in Chapter 4). In perception, there was a clear difference in the explicit goodness rating task, in which deletions were more penalized for /h/ than /ʔ/ (Experiment 3b). In spoken-word recognition, by contrast, there was a numeric difference between the two target sounds, but not significantly so (Experiment 3a). Interestingly, despite better discrimination in the explicit task (Experiment 2b), sound substitutions in the implicit task did not hinder learners in spoken-word recognition, and were hardly ever noticed by the learners (Experiment 2a). These results are in line with findings that learners perform typically better at explicit than implicit tasks (e.g., Llompart & Reinisch, in press_b; Sebastián-Gallés & Díaz, 2012). Moreover, the effect of explicit knowledge is larger in explicit tasks than in implicit, more natural listening situations, but it was not predominantly present in all cases. Given the a-priori assumption that learners may have a huge advantage for /h/ compared to a sound they are not even aware of, the findings lead to an interesting conclusion. This is, explicit knowledge of one sound within a novel sound pair has an influence on how the sounds are represented in the mental lexicon, but this effect was not as massive as expected. Rather, learners appear to be able to acquire L2 sounds without explicit instruction and extensive training, even without having a notion of the existence of this sound. This may be true, even when they start learning their second language later in life (see also, e.g., Flege, 1995).

Note that the glottal stop or glottalization is used in a different function in Italian (Bertinetto & Loporcaro, 2005; Stevens, Hajek, & Absalom, 2002). The comparably good performance of producing this sound may thus have been facilitated by activation of motor routines that are already in place. However, it is presumably not directly transferred from the production of glottal stops as onsets of vowel-initial words in L1-Italian, since those were produced in most of the cases without glottal stop or glottalization (Experiment 1). If the relatively high performance in the production of /ʔ/ is not directly transferred from L1, and orthography implies that these words should start with a vowel, learners must have caught up this sound from another source. This is likely exposure to native speech. Indeed, in contrast to the German learners of English in Chapters 2 and 3, all Italian learners of German in Chapter 4 had their permanent or temporary residence in a German-speaking environment and heard native German on a daily basis. Moreover, all Italian participants had started to learn German only at school or even later. This may therefore be helpful for all learners who do not have the opportunity to start learning a foreign language early in life. For learning languages in classrooms, explicit information to guide learners' attention towards contrasts can be helpful, but it is the teacher's task to evaluate how applicable this

information can be for learners, so that they can improve their speaking and perception skills to communicate in a L2.

To conclude, the findings of this thesis suggest that learners' representations of L2 words are not only non-target-like due to difficulties in perceiving unfamiliar sounds. Rather, in addition to that, representations are shaped by the accent typical of the learners' L1 (Chapter 2) and by their very specific production patterns in an L2 (Chapter 3). Furthermore, explicit knowledge can impact how well novel sounds are acquired, even though listeners are able to directly pick up cues from the input, which allows them to acquire sounds even without being aware of them (Chapter 4). The present thesis hence gives some insights in how representations of L2 words are shaped by explicit knowledge, familiarity with native and nonnative input, and one's own voice.

ZUSAMMENFASSUNG

Die vorliegende Arbeit beschäftigte sich mit der Beobachtung, dass viele Lerner*innen einer Fremdsprache (L2) mit einem hörbaren Akzent sprechen. Dies kann auch dann der Fall sein, wenn sie früh mit dem Fremdspracherwerb begonnen haben und selbst wenn sie die Fremdsprache häufig nutzen. Ein fremdsprachlicher Akzent ist vor allem darauf zurückzuführen, dass die Produktion und Perzeption von L2-Lauten von dem phonetisch-phonologischen System der Erstsprache (L1) beeinflusst wird (Best & Tyler, 2007; Flege, 2003). Dies hat zur Folge, dass je nach L1-L2 Kombinationen verschiedene Laute unterschiedlich schwer zu erwerben sind (z.B. van Leussen & Escudero, 2015). Aber auch innerhalb einer Lerner-Gruppe mit dem gleichen L1-Hintergrund gibt es Unterschiede in der Stärke des Akzents. Ein Faktor ist beispielsweise das Alter, in dem man mit dem Fremdspracherwerb beginnt (z.B. Piske, MacKay, & Flege, 2001). Die Experimente der vorliegenden Arbeit widmeten sich dem Phänomen des fremdsprachlichen Akzentes mit einem etwas anderen Fokus: Wie kommt es, dass es so schwierig erscheint, seinen Akzent zu verlieren? Die Arbeit betrachtete dabei verschiedene Faktoren: Vertrautheit und Erfahrung mit dem Akzent der eigenen Erstsprache (Kapitel 2), die Vertrautheit mit eigenen, ganz persönlichen Sprechmustern in einer Fremdsprache (Kapitel 3), und das Bewusstsein über die Existenz schwieriger Laute in der Fremdsprache (Kapitel 4).

Dabei wurden zwei unterschiedliche Arten von fremdsprachlichen Kontrasten untersucht: Einerseits Lautkontraste, die schwierig sind, da zwei L2-Kategorien in artikulatorischer und akustischer Hinsicht einer einzigen L1-Kategorie ähnlich sind. Dies ist beispielsweise bei dem englischen Vokalkontrast /ɛ-æ/ wie in *bet-bat* („wetten“ - „Fledermaus“) für Lerner*innen mit Deutsch als L1 der Fall. Im Deutschen gibt es im vorderen halboffenen Vokalraum nur /ɛ/ als ungerundeten Vokal. Daher wird dieser häufig für beide englischen Kategorien produziert, auch wenn das Wort eigentlich das etwas offenere /æ/ enthalten sollte, wie in *„h[ɛ]ppy“ (Llompart & Reinisch, in press_b). Ein ähnlicher Fall ist der Stimmhaftigkeitskontrast in wortfinalen Obstruenten, wie *feet-feed* („Füße“ - „füttern“) oder *face-phase* („Gesicht“ - „Phase“). Im Deutschen gibt es einen Stimmhaftigkeitskontrast in Obstruenten, dieser wird aber in wortfinaler Position minimiert und nur der stimmlose/fortis Obstruent produziert („Auslautverhärtung“, siehe aber auch Kleber, John, & Harrington, 2010; Roettger, Winter, Grawunder, Kirby, & Grice, 2014). Dieser Prozess wird von Lerner*innen oft ins L2-Englische übertragen, wodurch

Wörter wie *feet* und *feed* sehr ähnlich klingen, wenn sie von Deutschen gesprochen werden. Ein fremdsprachlicher Akzent zeigt sich also oft darin, dass unbekannte Laute der L2 nicht differenziert genug ausgesprochen und somit relevante Kontraste akustisch minimiert werden. Daher können Wörter, die mit einem Akzent gesprochen werden, häufig schwieriger zu verstehen sein als solche ohne fremdsprachlichen Akzent. Der Akzent ist dabei typischerweise in jenen Lauten stärker, die nicht aus der L1 bekannt sind. Daher werden der Vokal /ɛ/ wie in *bet* und stimmlose wortfinale Obstruenten wie in *feet* und *face* in dieser Arbeit als einfache Laute bezeichnet, wohingegen /æ/ wie in *bat* und phonemisch stimmhafte Obstruenten wie in *feed* und *phase* als schwierige Laute gelten. Die Produktion und Perzeption dieser Kontraste, in denen jeweils ein Laut aus der L1 bekannt ist und der andere nicht, wurden in den Kapiteln 2 und 3 mit deutschen Lerner*innen des Englischen untersucht.

Eine andere Art von L2-Lauten, die Lerner*innen Schwierigkeiten bereiten können, sind solche Laute, die artikulatorisch und auditiv sehr unterschiedlich zu jeglichen L1-Kategorien sind. Ein Beispiel hierfür ist der glottale Frikativ /h/ im Deutschen oder Englischen für Lerner*innen mit Italienisch als L1, wo dieser Laut nicht kontrastiv verwendet wird (Krämer, 2009). Im Deutschen kontrastiert der glottale Frikativ mit dem Glottalverschluss /ʔ/, der kanonisch vor wort-initialen Vokalen als Verschluss oder Glottalisierung produziert wird (Kohler, 1994). Diese beiden Segmente existieren zwar im Italienischen, aber nur in paralinguistischer Funktion oder zur Signalisierung von Betonung (Bertinetto & Loporcaro, 2005; Stevens, Hajek, & Absalom, 2002). Im Gegensatz zu den oben genannten Lautkontrasten gibt es in diesem Fall keine offensichtliche L1-Kategorie, durch welche der schwierige Laut ersetzt werden könnte. Die Experimente in Kapitel 4 untersuchten daher, wie italienische Lerner*innen des Deutschen diese beiden Laute, /h/ und /ʔ/, produzieren und wahrnehmen.

Wie bereits erwähnt wirkt sich ein fremdsprachlicher Akzent häufig auf die Verständlichkeit von Wörtern aus. Um Wörter produzieren und verstehen zu können, müssen nicht nur einzelne Laute wahrgenommen werden, sondern auch Wörter mit diesen Lauten in einem mentalen Lexikon dargestellt und zugänglich sein. Ein Ziel der vorliegenden Arbeit war es daher auch, etwas darüber zu erfahren, wie L2-Wörter im mentalen Lexikon dargestellt sein könnten. Dafür wurden Methoden verwendet, die Rückschlüsse auf diese Repräsentationen zulassen können.

Das Experiment in Kapitel 2 ging der Frage nach, ob L2-Wörter im mentalen Lexikon typische Akzentmerkmale der L1 tragen. Dafür wurde untersucht, wie deutsche

Englischlerner*innen die Aussprache von englischen Wörtern mit deutschem Akzent bewerten. Dass Repräsentationen von L2-Wörtern im mentalen Lexikon vom Akzent der eigenen L1 gefärbt sind, z.B. englische Wörter mit deutschem Akzent, wäre aus folgenden Gründen denkbar: Die Repräsentationen sind vermutlich von den phonetisch-phonologischen Eigenschaften der L1 beeinflusst, bzw. durch die Unterschiede zwischen L1 und L2 (z.B. Cutler, 2015). Dies könnte noch dadurch verstärkt werden, dass L2-Lerner*innen häufig untereinander Kontakt haben und viele Produktionen mit Akzentmerkmalen hören, vor allem in typischen Sprachkurs-Situationen. Wenn es der Fall ist, dass sich Erfahrung und Vertrautheit mit diesem Akzent in den mentalen Repräsentationen von L2-Wörtern widerspiegelt, dann sollten Lerner*innen solche Wörter auch als akzeptabel empfinden. Des Weiteren sollten den Lerner*innen Akzentmuster ihrer eigenen L1 umso eher auffallen, je mehr Kontakt sie zu Äußerungen von englischen Muttersprachler*innen haben. Das würde darauf hindeuten, dass Wortrepräsentationen im mentalen Lexikon durch Erfahrung besser werden.

Diese Fragestellungen wurden in einem Wahrnehmungsexperiment mit deutschen Englischlerner*innen und Muttersprachler*innen des Englischen als Kontrollgruppe untersucht, welche die Aussprache verschiedener Wörter zu beurteilen hatten. Als Material wurden Produktionen von einer anderen Gruppe deutscher Englischlernerinnen verwendet. Die gesprochenen Wörter bildeten Minimalpaare mit den oben genannten Kontrasten: der Vokalkontrast /ɛ-æ/ und die Stimmhaftigkeitskontraste. Die zu bewertenden Wörter unterschieden sich also darin, ob sie einen einfachen oder schwierigen Laut enthielten, wobei der Akzent typischerweise in letzteren stärker ist. Die Stärke des Akzentes konnte daher akustisch gemessen werden: Je kleiner der produzierte akustische Unterschied zwischen den Ziellauten (z.B. zwischen /ɛ/ und /æ/) war, desto stärker sollte der Akzent sein, typischerweise in dem schwierigen Laut. Anhand dieses produzierten Kontrastes wurde das Material zu einem von drei Materialsets zugeordnet (gut, mittel, schlecht). „Gut“ bedeutet hier, dass der akustische Unterschied zwischen den präsentierten Wörtern groß war, und „schlecht“ bedeutet, dass der Kontrast minimiert war, und beide Wörter der Minimalpaare sehr ähnlich klangen. Stimuli des Materialsets „mittel“ lagen hinsichtlich des Kontrastes zwischen den beiden anderen. Die Aufgabe der Proband*innen war es zu bewerten, wie gut diese Wörter ausgesprochen wurden.

Zwei Haupterkenntnisse konnten aus den Ergebnissen gezogen werden. Erstens, je mehr Erfahrung die Lerner*innen mit gesprochenem Englisch hatten, desto eher hörten sie die Unterschiede im Akzent, ähnlich wie die Muttersprachler*innen in der Kontrollgruppe.

Zweitens, die Lerner*innen, deren Kontakt mit der Fremdsprache eher auf Englisch mit deutschen Akzentmerkmalen beschränkt war (beispielweise im Kontakt zu anderen Lerner*innen), bewerteten die Wörter kaum unterschiedlich, was darauf schließen lässt, dass sie die Unterschiede im Akzent weniger wahrnahmen. Im Gegensatz zu früheren Studien (Flege, Munro, & MacKay, 1995; Thompson, 1991) war die Aufgabe nicht, den Akzent direkt zu beurteilen, sondern wie gut die Wörter ausgesprochen wurden. Bei der Beurteilung, wie gut ein Wort ausgesprochen wird, muss es mit einer vorhandenen Repräsentation verglichen werden. Wenn ein gehörtes Wort als gut beurteilt wird, ist das ein Hinweis darauf, dass eine Übereinstimmung mit einer Repräsentation im mentalen Lexikon gefunden wurde. In diesem Sinne deuten die Ergebnisse darauf hin, dass L2-Wörter im mentalen Lexikon Charakteristika des Akzentes der eigenen L1 tragen. Diese Repräsentationen können sich aber durch Kontakt zu der Fremdsprache weiterentwickeln. Den Akzent der eigenen L1 in einer Fremdsprache zu erkennen, könnte ein maßgeblicher Punkt sein, die Aussprache zu verbessern. Den Akzent nicht wahrzunehmen, im Gegenteil, ist ein möglicher Faktor, der eine Verbesserung verhindert.

Die Experimente in Kapitel 3 erweiterten die Untersuchung von Vertrautheit mit fremdsprachlichen Aussprachemustern. Die Fragestellung in diesem Kapitel war, ob Lerner*innen besser darin sind, *eigene* L2-Produktionen als das intendierte Wort zu verstehen als Produktionen von anderen, welche die Wörter ähnlich gut produzierten (d.h. die Wörter in der Aussprache ähnlich gut differenzierten). Dafür wurden dieselben englischen Minimalpaare wie in Kapitel 2 verwendet. Sprecherinnen des Englischen mit Deutsch als L1 produzierten die Wörter der Minimalkontraste einzeln in randomisierter Reihenfolge. Anhand des produzierten akustischen Unterschiedes zwischen den Wörtern der Minimalpaare (z.B. *b**e**t-b**a**t*) wurden die Sprecherinnen pro Kontrast in eine von drei Gruppen eingeteilt (gut, mittel, schlecht). Wie oben bedeutet „gut“, dass der Kontrast innerhalb der Probandengruppe am größten produziert wurde, und „schlecht“, dass die Wörter innerhalb des Kontrastes akustisch nicht oder nur minimal differenziert wurden. Sprecherinnen, die der „mittleren“ Gruppe zugeteilt waren, lagen in den produzierten Kontrasten zwischen den beiden anderen.

In zwei Perzeptionsexperimenten war es die Aufgabe der Teilnehmerinnen, Wörter aus diesen Minimalpaaren zu identifizieren. In Experiment 1 wurden den Teilnehmerinnen Wörter, die sie selbst gesprochen hatten, und Aufnahmen von anderen vorgespielt, welche die Kontraste ähnlich gut produziert hatten (d.h. derselben Produktionsgruppe zugeordnet waren). In Experiment 2 mussten die Teilnehmerinnen der

guten und schlechten Gruppe Aufnahmen der jeweils anderen Gruppe verstehen (d.h. Teilnehmerinnen aus der guten Gruppe hörten Aufnahmen der schlechten Produktionsgruppe und umgekehrt).

Experiment 1 ergab, dass Lernerinnen tatsächlich besser darin waren, ihre eigenen Produktionen zu verstehen als die der anderen, obwohl diese die Kontraste ähnlich gut produziert hatten. Dieser Vorteil für die eigenen Produktionen war ähnlich stark in allen Produktionsgruppen. Die Erkenntnisse aus Experiment 2 waren, dass gute Produktionen (d.h. deutlichere akustische Unterschiede bei der Produktion der Kontraste) insgesamt von allen Lernerinnen besser verstanden wurden. Sprecherinnen, welche die Kontraste selbst besser produziert hatten, profitierten aber zu einem größeren Ausmaß. Ein zusätzlicher Vergleich mit den Daten aus Experiment 1 ergab, dass ein gut produzierter akustischer Kontrast insgesamt wichtiger für die Wortidentifizierung war, als seine eigenen Produktionen zu hören. Wenn die akustischen Hinweise im Signal aber ausreichend waren und die eigene Stimme gehört wurde, dann wurde die Worterkennung dadurch noch zusätzlich verbessert. Diese Ergebnisse lassen darauf schließen, dass Repräsentationen von L2-Wörtern zusätzlich zu allgemeinen sprachlichen Akzentmustern typisch für die L1 auch noch von den eigenen, ganz persönlichen Sprechmustern in einer L2 geformt werden. Eine erhöhte Vertrautheit mit persönlichen Sprechmustern könnte sich aber auch darauf auswirken, wie der eigene Akzent wahrgenommen wird, der sich ja in den akustischen Eigenschaften der produzierten Wörter widerspiegelt. Kontrollmechanismen beim Sprechen vergleichen den wahrgenommenen Input mit dem geplanten Output (z.B. Tourville & Guenther, 2011). Wenn die eigene Sprache mit Repräsentationen verglichen wird, die von ständigem Hören des eigenen Akzentes geformt sind, könnte eine Diskrepanz nicht wahrgenommen werden, und somit auch kein Grund, seine Produktionen anzupassen. In anderen Worten, besseres Verständnis der eigenen Produktionen und ein vermindertes Bewusstsein des eigenen Akzentes könnten zwei Seiten derselben Medaille sein.

Die Ergebnisse aus den vorigen Kapiteln deuten darauf hin, dass Lerner*innen sich an den Akzent der eigenen Erstsprache allgemein (Kapitel 2) und der spezifischen persönlichen Produktionsmuster (Kapitel 3) gewöhnen, und dass sich dies im mentalen Lexikon widerspiegelt. Diese Anpassung an den Akzent ist vermutlich das Resultat eines langen und unbewussten Prozesses. Allerdings gibt es auch Faktoren in einer L2, derer sich Lerner*innen bewusst sind, beispielsweise Orthographie. Diese kann unter Anderem suggerieren, dass zwei lautliche Darstellungen unterschiedlich ausgesprochen werden (selbst wenn sie gar nicht kontrastiert werden, Bassetti, Sokolović-Perović, Mairano, &

Cerni, 2018). Orthographie könnte auch als Hinweis darauf verwendet werden, ob ein Laut existiert oder nicht. Die Experimente in Kapitel 4 untersuchten daher, ob explizites Wissen über die Existenz von L2-Lauten einen Einfluss darauf hat, wie diese Laute im mentalen Lexikon dargestellt sind, und wie gut sie in Produktion und Perzeption genutzt werden.

Dafür wurden Lerner*innen des Deutschen mit Italienisch als L1 getestet. Der Fokus lag auf dem deutschen Lautpaar /h-ʔ/, das nur fuß-initial vor Vokalen, wie in [h]ut oder [ʔ]apfel vorkommt. Beide Laute sind im Deutschen gleich distribuiert und beide werden von Muttersprachler*innen für die Worterkennung genutzt (Mitterer & Reinisch, 2015). Jedoch unterscheiden sie sich maßgeblich in ihrem Status: Während /h/ orthographisch kodiert und eine bekannte Schwierigkeit für italienische Lerner ist, wird /ʔ/ orthographisch nicht dargestellt, und typischerweise sind sich weder Muttersprachler*innen noch Lerner*innen dieses Lautes bewusst. Wenn explizites Wissen einen Einfluss auf den Erwerb von L2-Lauten hat, dann müssten italienische Lerner*innen /h/ besser erlernen als /ʔ/. Um dies zu testen, wurden ein Produktions- und mehrere Perzeptionsexperimente durchgeführt. In Experiment 1 wurden Wörter mit den Ziellauten von italienischen Lerner*innen und einer deutschen Kontrollgruppe produziert. Die Perzeptionsexperimente testeten, wie Lerner*innen Lautsubstituierungen (Experiment 2, [ʔ]ut für *Hut* und [h]apfel für *Apfel*) und Lautelisionen wahrnehmen (Experiment 3). Für beide der Bedingungen gab es eine explizite und eine implizite Aufgabe, da gezeigt wurde, dass Lerner*innen gut darin sein können, L2-Kontraste zu unterscheiden, wenn sie explizit danach gefragt werden. Diese Fähigkeit überträgt sich aber nicht zwingend darauf, wie gut sie Wörter mit diesen Kontrasten in natürlicheren Situationen erkennen und produzieren (Llompart & Reinisch, in press_b; Sebastián-Gallés & Díaz, 2012). In der expliziten Aufgabe musste die Aussprache der Wörter in den jeweiligen Bedingungen (substituiert oder elidiert) beurteilt werden. Die implizite Aufgabe verwendete die Eye-Tracking-Methode. Dabei werden die Blicke der Teilnehmer*innen zu Wörtern oder Bildern auf einem Bildschirm gemessen (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974). Diese Blickmuster können Hinweise auf die Form von Einträgen im mentalen Lexikon geben, da angenommen wird, dass bei der Worterkennung sprachlicher Input mit vorhandenen lexikalischen Einträgen verglichen wird (Huettig, Rommers, & Meyer, 2011). In keinem der Experimente wurden geschriebene Wörter sondern nur Bilder gezeigt, um den Einfluss von explizitem Wissen auf den Lauterwerb und nicht einen direkten Effekt der Orthographie zu testen.

Die Ergebnisse zeigten, dass Lerner*innen beide Laute in etwa 70% korrekt produzierten. Beide Laute wurden gleich oft substituiert, doch /ʔ/ wurde häufiger elidiert. Die Perzeptionsexperimente ergaben, dass Lerner*innen die Laute in den expliziten Aufgaben insgesamt unterscheiden konnten, was sich an schlechteren Bewertungen für Wörter mit dem substituierten Laut widerspiegelte (Experiment 2). Auch Wörter, in denen die Laute elidiert wurden, erhielten schlechtere Bewertungen als die korrekten Formen, doch der Unterschied war bei /h/ größer als bei /ʔ/ (Experiment 3). Die Eye-Tracking-Experimente ergaben, dass Lerner*innen die Laute zur Erkennung gesprochener Wörter nutzten, vor allem für /h/-initiale Wörter (Experiment 3), aber welcher Laut vorhanden war, machte keinen Unterschied (Experiment 2). Das lässt darauf schließen, dass Lerner*innen eine einzelne Kategorie gebildet haben, die beide glottalen Laute umfasst, in welcher /h/ dominanter zu sein scheint. Die Unterschiede zwischen den Ziellauten deuten darauf hin, dass explizites Wissen über einen neuen Laut dazu beitragen kann, diesen zu erlernen, wobei der Vorteil deutlicher in expliziten Aufgaben ist. Die Ergebnisse deuten aber auch darauf hin, dass L2-Laute erlernt werden können, selbst wenn Lerner*innen sich der Existenz dieser Laute nicht einmal bewusst sind. Diese Ergebnisse tragen dazu bei, den Erwerb von jenen L2-Lauten besser zu verstehen, die keiner L1-Kategorie ähnlich sind.

Insgesamt können aus den Ergebnissen zwei Hauptschlussfolgerungen gezogen werden. Die erste ist, dass Repräsentationen von fremdsprachlichen Wörtern nicht einfach unpräzise sind. Vielmehr deuten die Ergebnisse darauf hin, dass L2-Wörter im mentalen Lexikon zusätzlich von allgemeinen Akzentmerkmalen der L1 (Kapitel 2) als auch von ganz persönlichen Sprechmustern (Kapitel 3) geprägt sind. Das bedeutet, durch vermehrtes Hören typischer Akzentmuster wird Akzent im mentalen Lexikon noch verstärkt. Dies kann aber zur Folge haben, dass Lerner*innen ihre typischen Akzentmuster, in denen sich ja die Diskrepanzen zwischen L1- und L2-Lauten widerspiegeln, nicht wahrnehmen, und somit nicht die Notwendigkeit bemerken, ihre Aussprache zu verändern. Die zweite Schlussfolgerung ist, dass explizites Wissen über einen schwierigen L2-Laut dabei helfen kann, diesen zu erwerben, aber nicht absolut notwendig ist. In anderen Worten, Lerner*innen können durch Kontakt zur L2 gesprochen von Muttersprachler*innen neue Laute erwerben, ohne sich überhaupt über deren Vorkommen bewusst zu sein. Dies ist eine vielversprechende Nachricht für Menschen, die eine Fremdsprache ohne explizite Instruktionen und später im Leben erwerben. Die Erkenntnisse dieser Arbeit können zum besseren Verständnis der Aussprache von L2-Lauten beitragen. Sie können aber auch Hinweise darauf geben, wie sich Lerner*innen in ihrer Produktion und der Wahrnehmung

von L2-Lauten und -Wörtern verbessern, was ihnen eine bessere Kommunikation in der Fremdsprache ermöglichen könnte.

Appendix I

Appendix I contains additional information on the materials and acoustic measures of the stimuli used in the experiment in Chapter 2.

Appendix I.A: Materials

Table I.A: Words and word pairs that were recorded in the production session. In the minimal pairs, the word after the dash is the one containing the critical difficult sound. All words used in the experiment are monosyllabic. The words in italics were recorded and acoustically analyzed but excluded from the materials for the perception experiment. The filler words were recorded to distract the speakers from the purpose of the study but they were not further analyzed.

Vowels /ɛ/ – /æ/	Fricatives	Stops	Filler words	
bet – bat	ice – eyes	feet – feed	car	piece
flesh – flash	leaf – leave	pick – pig	eat	shine
men – man	race – raise	root – rude	fine	ship
pen – pan	rice – rise	rope – robe	force	sing
set – sat	safe – save	sight – side	fourth	sit
			forest	skin
bed – bad	*face – phase*	*back – bag*	get	state
dead – Dad	*proof – prove*	*bat – bad*	honest	strong
head – had		*bet – bed*	king	time
letter – latter		*bright – bride*	kiss	worse
merry – marry		*brought– broad*	nine	worth
send – sand		*heart – hard*		
		height – hide		
		white – wide		

Appendix I.B: Acoustic measures of the stimuli for the goodness rating task
The Figures below show a selection of acoustic measures that had been used to determine the produced difference between the words of the minimal pairs. Tokens were assigned to material sets A, B, or C for each type of contrast according to different acoustic measures (see text for details). The variability in the boxes is due to inter-speaker differences (4 speakers per group) and to the different words (5 words per category and contrast).

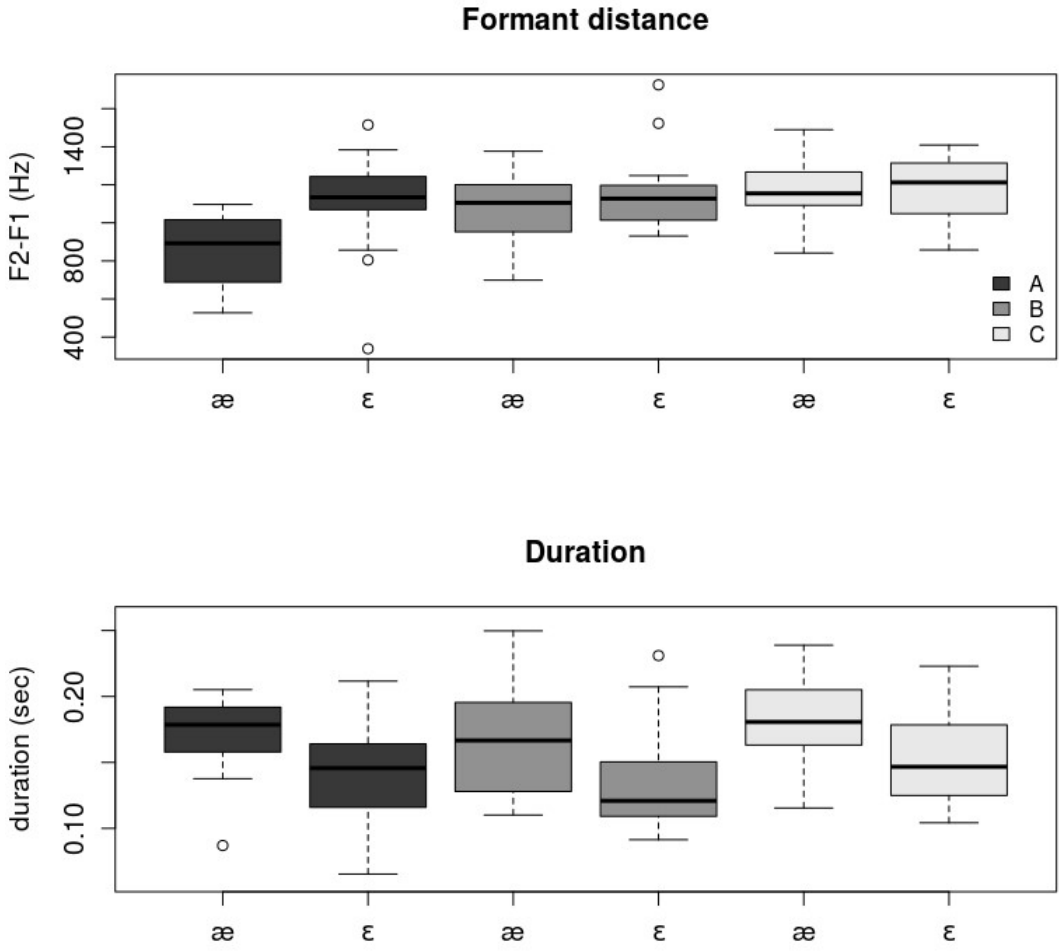

Figure I.B.1 Upper Panel: Formant values measured as the difference between F2 and F1 in Hz during a stable segment in the vowel for words with either /æ/ or /ɛ/ for the German learners grouped into three groups of 4 (dark grey = group A, mid grey = group B, light grey = group C); Lower Panel: Duration values of the entire vowel for words with either /æ/ or /ɛ/ and the different groups.

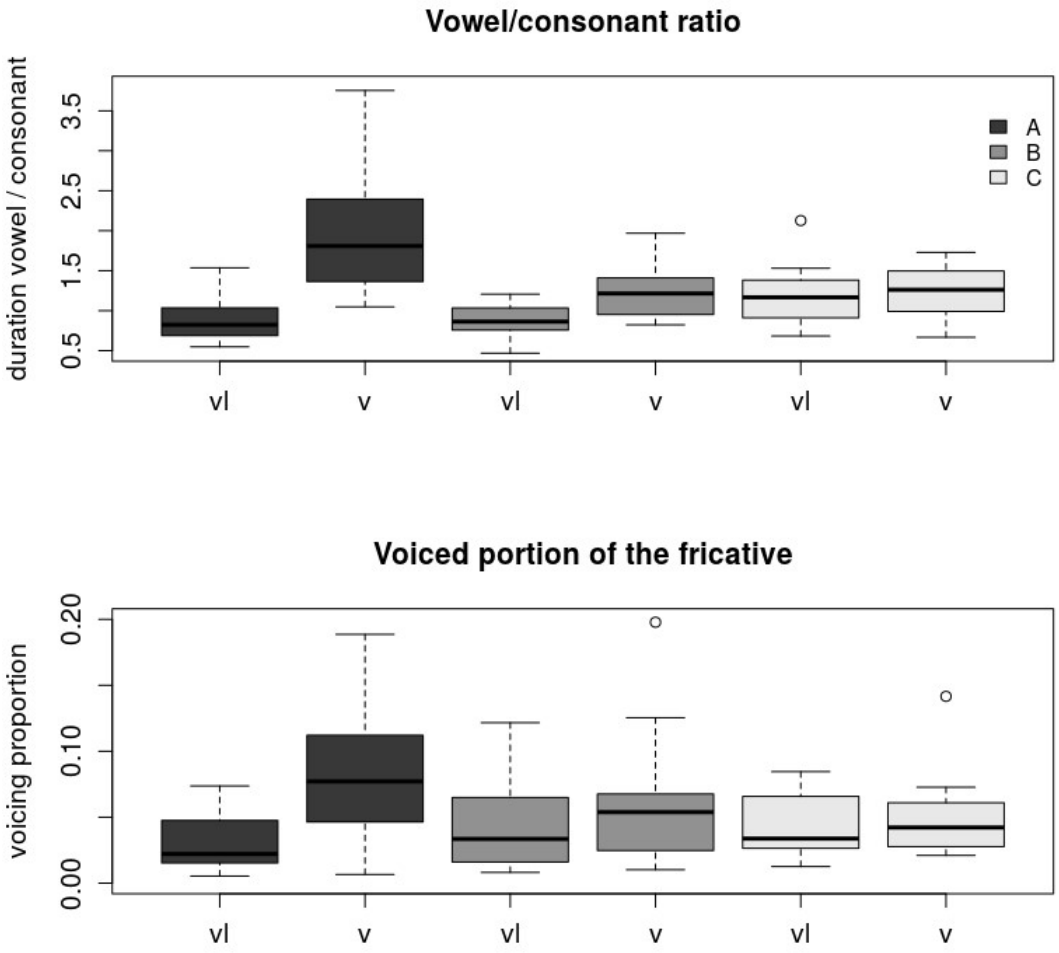

Figure I.B.2 Upper panel: Vowel/consonant ratios measured as the duration of the vowel divided by the duration of the consonant in words ending in voiced (v) or voiceless (vl) fricatives for the German learners grouped into three groups of 4 (dark grey = group A, mid grey = group B, light grey = group C); Lower Panel: Voiced portion of the fricative measured as the duration of the voiced part of the fricative divided by the total duration.

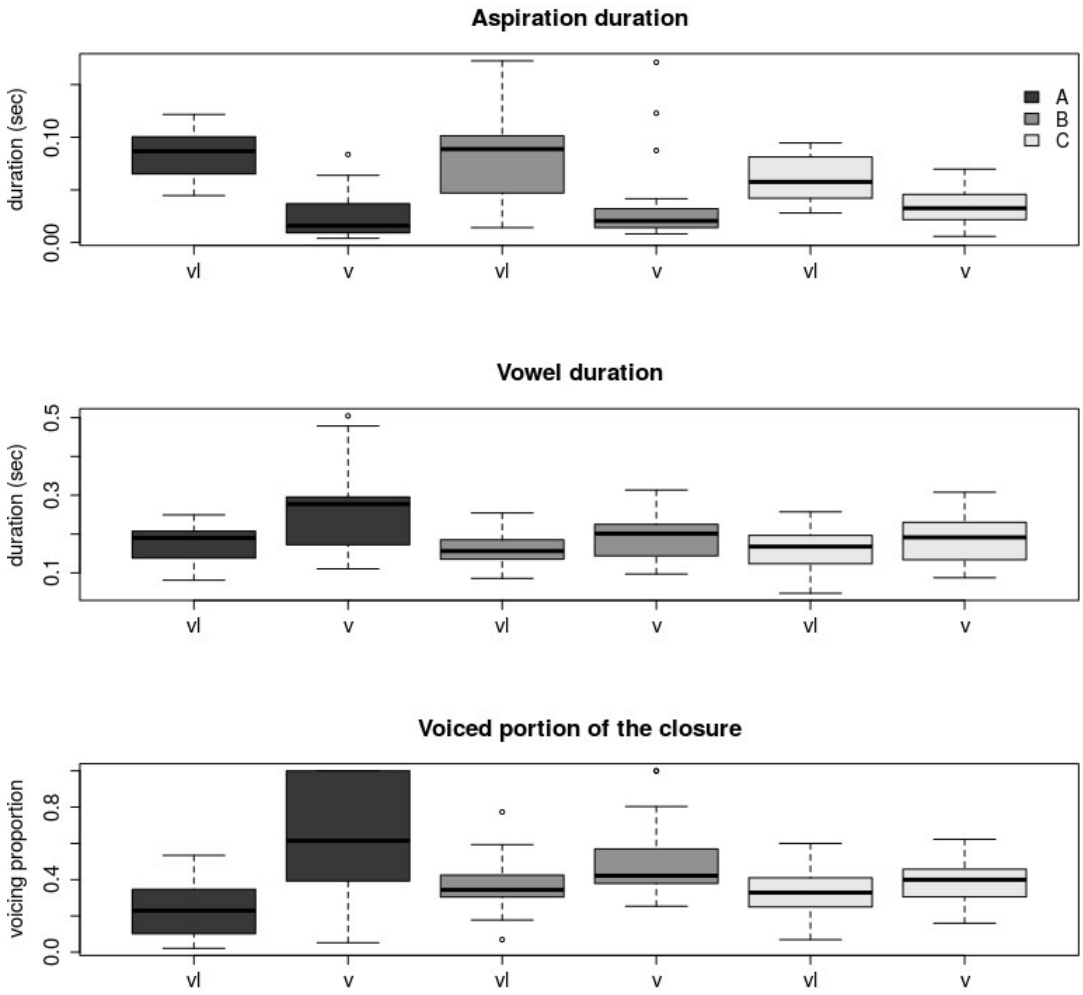

Figure I.B.3 Top panel: Aspiration duration for words ending in either voiced (v) or voiceless (vl) stops for the German learners grouped into three groups of 4 (dark grey = group A, mid grey = group B, light grey = group C); Mid panel: Duration of the preceding vowel; Bottom panel: Voiced portion of the closure measured as the duration of the voicing during closure divided by the total closure duration. As all other words, words containing a word-final stop were embedded in the end of carrier sentences. All word-final stops were produced as released stops.

Appendix II

Appendix II contains additional information on the materials and the acoustic measures of the productions in the experiments reported in Chapter 3.

Appendix II.A: Materials

Table II.A.1: Words and word pairs that were recorded in the production session. For the minimal pairs, the word after the dash is the one containing the critical difficult sound. The starred words and their counterparts were recorded and analyzed but excluded from the materials for the perception experiment because several speakers pronounced the vowels in "broad", "height" and "prove" differently than in the other word of the pair. The word "latter" and its counterpart were excluded because several participants reported that they did not know the meaning of the word.

Vowels /ɛ/ – /æ/	Fricatives	Stops	Filler words	
bed –bad	ice – eyes	back – bag	car	piece
bet – bat	face – phase	bat – bad	eat	shine
dead – Dad	leaf – leave	bet – bed	fine	ship
flesh – flash	proof – prove*	bright – bride	force	sing
head – had	race – raise	brought – broad*	fourth	sit
letter –	rice – rise	feet – feed	forest	skin
latter**	safe – save	heart – hard	get	state
men – man		height* – hide	honest	strong
merry – marry		pick – pig	king	time
pen – pan		root – rude	kiss	worse
send – sand		rope – robe	nine	worth
set – sat		sight – side		
		white – wide		

Table II.A.2: Carrier sentences used in the production task. Sentences and words were randomly paired for each participant.

Number	Sentence
1	Here is the word
2	She forgot the word
3	He knows the word
4	You say the word
5	You read the word
6	The next word is
7	The correct term is
8	The next term is
9	The next expression is
10	The right expression is

APPENDIX II.B: Acoustic measures

The Figures below show a selection of acoustic measures that had been taken to determine the produced difference between the words of the minimal pairs for each of the three sound contrasts. These measures were used to assign participants to proficiency groups. For the vowels, the first two formants and duration were measured. For the word-final fricatives, the duration of the preceding vowel and the fricative were combined to the ratio between vowel duration and fricative duration. In addition, the voiced portion of the fricative was measured. For the word-final stops, the duration of the aspiration, the duration of the preceding vowel and the voiced portion of the closure were taken into account. Cues to each contrast were weighted in the order named above.

The speakers were assigned one by one to the proficiency groups A, B, and C. Since the whole group of participants had to be distributed, they were split into three groups of 8 speakers each. First, the eight speakers with the clearest contrasts, according to the cues listed above, were assigned to group A. Then, the eight speakers with the smallest contrasts were selected for group C. The remaining eight participants were assigned to group B (see also Method section). In the Figures below, the acoustic measures are averaged over the two repetitions and words. The variability is hence due to inter-speaker differences (8 speakers per box).

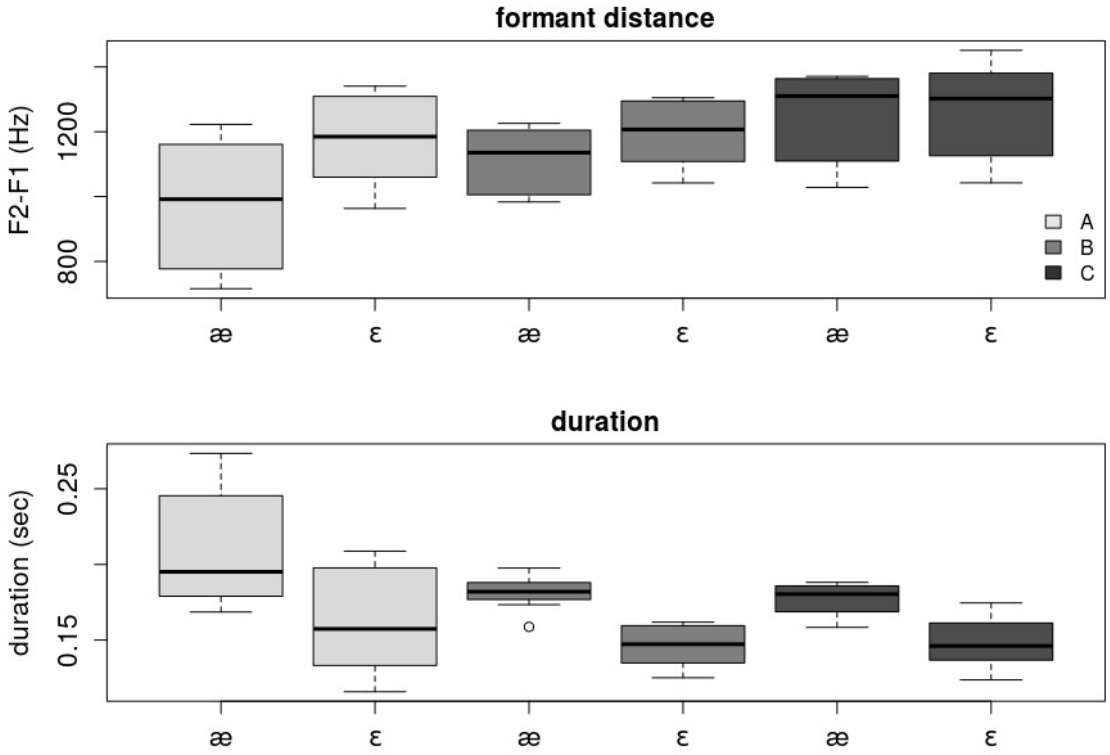

Figure II.B.1 Upper Panel: Formant values measured as the difference between F2 and F1 in Hz during a stable segment in the vowel for words with either /æ/ or /ɛ/ for the German learners grouped into three groups of 8 (light grey = group A, mid grey = group B, dark grey = group C); Lower Panel: Duration values of the entire vowel for words with either /æ/ or /ɛ/ and the different groups.

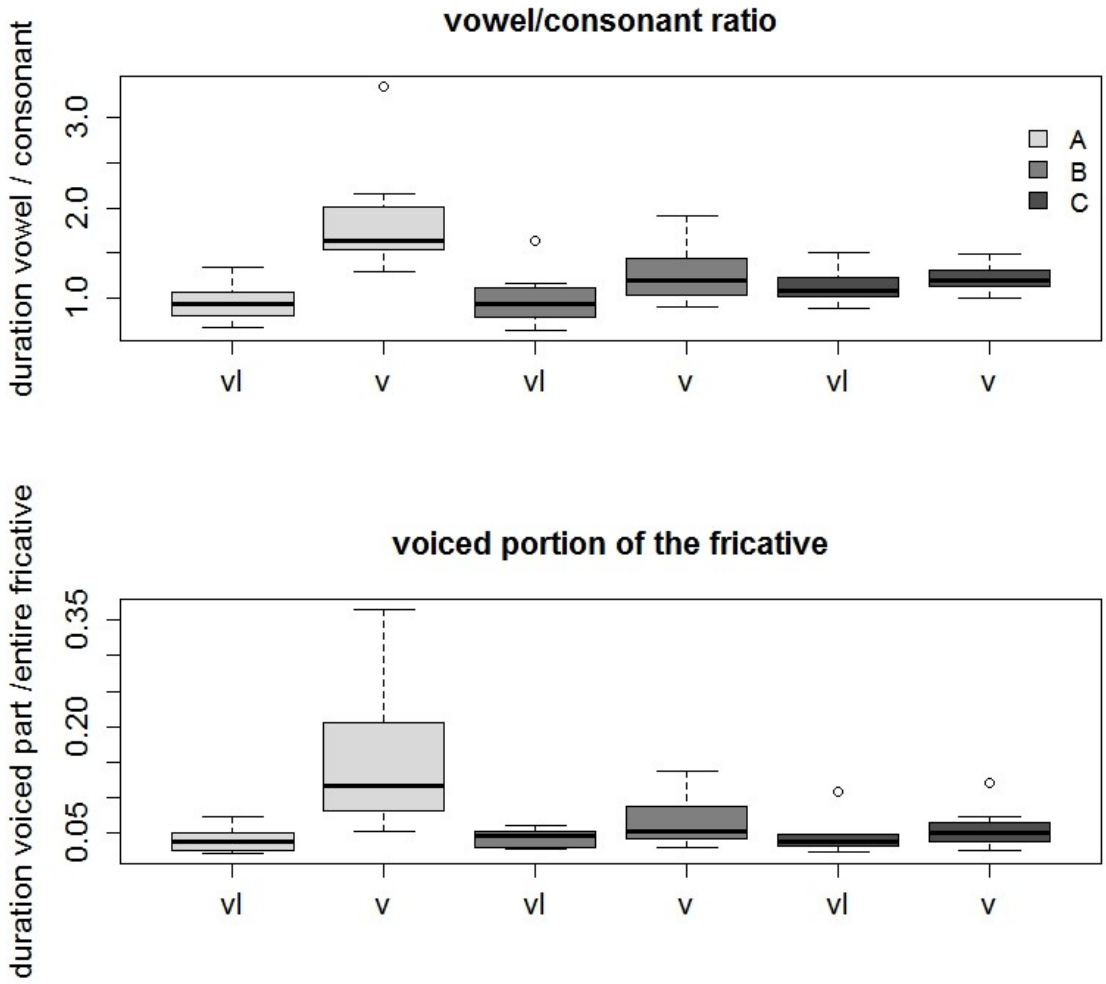

Figure II.B.2 Upper panel: Vowel/consonant ratios measured as the duration of the vowel divided by the duration of the consonant in words ending in voiced (v) or voiceless (vl) fricatives for the German learners grouped into three groups of 8 (light grey = group A, mid grey = group B, dark grey = group C); Lower Panel: Voiced portion of the fricative measured as the duration of the voiced part of the fricative divided by the total duration.

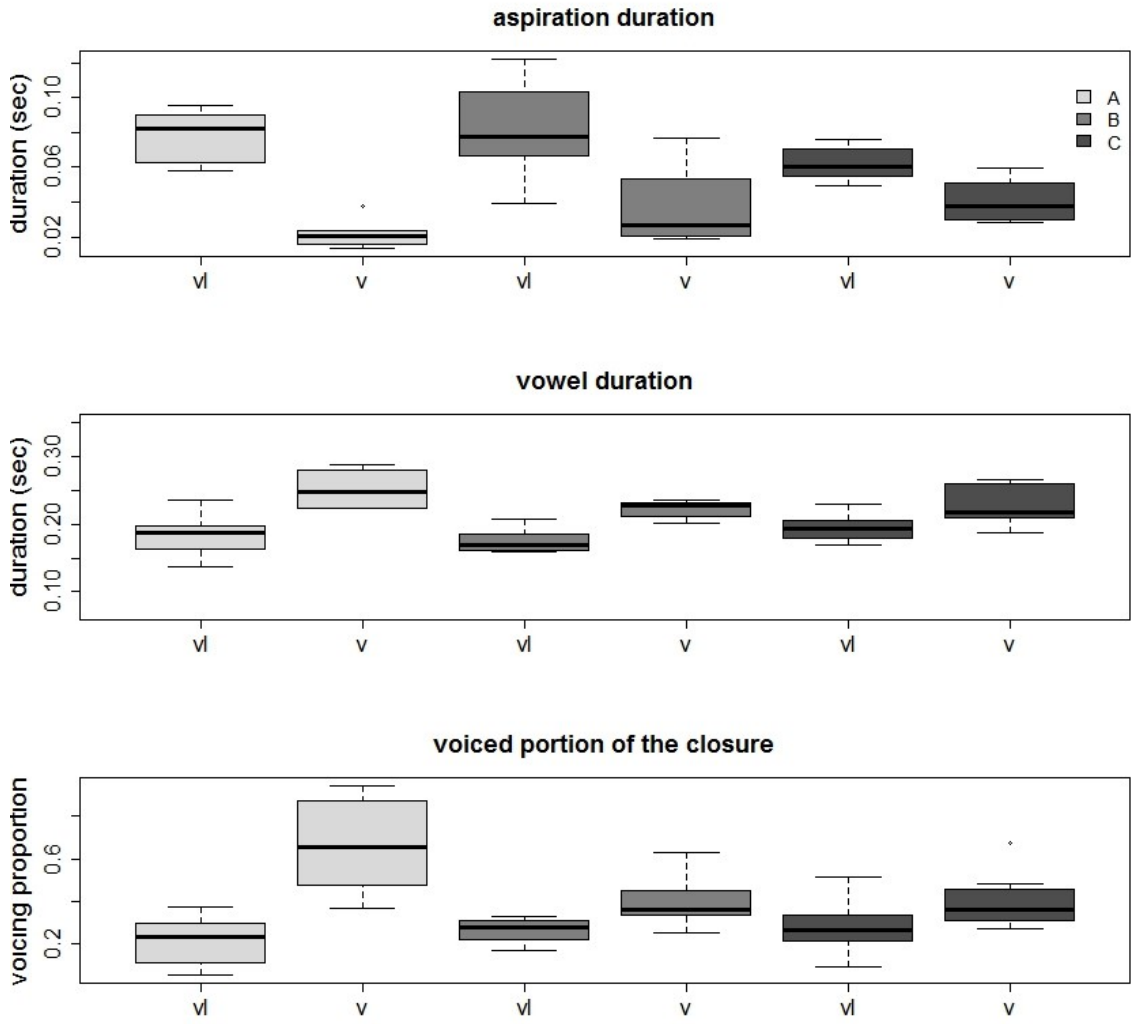

Figure II.B.3 Top panel: Aspiration duration for words ending in either voiced (v) or voiceless (vl) stops for the German learners grouped into three groups of 8 (light grey = group A, mid grey = group B, dark grey = group C); Mid panel: Duration of the preceding vowel; Bottom panel: Voiced portion of the closure measured as the duration of the voicing during closure divided by the total closure duration. As all other words, words containing a word-final stop were embedded in the end of carrier sentences. All word-final stops were produced as released stops.

Appendix III

Appendix III contains additional background information on the participants, information on the materials, and the results from the pretest in the experiments reported in Chapter 4.

Appendix III.A: Participant information from the questionnaire

Table III.A. Selected questions from the questionnaire: Length of residence in months (LoR), self-reported frequency of speaking (Spk frq) and listening (Lis frq) German, self-estimated proficiency in speaking (Spk prof) and listening (Lis prof) German, and self-estimated accent (Accent) in German. The values are reported in means and standard deviations (in brackets).

Group/task	LoR	Spk frq	Lis frq	Spk prof	Lis prof	Accent
Production	39.5(37.3)	2.8(1.4)	2.3(1.3)	3.8(1.3)	2.5(1.1)	5.2(1.2)
Perception 1	25.3(23.0)	2.7(1.6)	1.9(1.1)	3.4(1.0)	3(1.3)	5.4(1.3)
Perception 2	33.2(26.7)	2.9(1.6)	2.2(1.2)	3.8(1.3)	2.7(1.1)	5.6(1.1)

Appendix III.B: Materials

Table III.B.1. Target words used in the production task of the German sentences. These words were in plural and preceded by one of the numbers *neun*, *zehn* ("nine, ten") for the nasal, and *zwei*, *drei* ("two, three") for the vowel context.

/h/-initial words	/ʔ/-initial words
Hafen *harbor*	Adler *eagle*
Hähnchen *chicken*	Ampel *traffic light*
Hai *shark*	Apfel *apple*
Hamburger *burger*	Aprikose *apricot*
Hammer *hammer*	Arzt *doctor*
Hamster *hamster*	Ausgang *exit*
Hand *hand*	Auto *car*
Handschuh *glove*	Avokado *avocado*
Handtuch *towel*	Ei *egg*
Handy *mobile phone*	Elefant *elephant*
Hase *hare*	Engel *angel*
Haus *house*	Ente *duck*
Heft *booklet*	Erdbeere *strawberry*
Helm *helmet*	Esel *donkey*
Hemd *shirt*	Eule *owl*
Herz *heart*	Indianer *Indian*
Himbeere *raspberry*	Insel *island*
Hose *pants*	Ohrring *earring*
Hund *dog*	Olive *olive*
Hut *hat*	Uhr *watch*

Table III.B.2. Target words used in the production task of the Italian sentences. Only native speakers of Italian participated in this task. For the nasal context, target words of masculine gender in singular were preceded by the word *buon* ("good"). For the vowel context, target words of masculine or feminine gender in plural were preceded by a form of buono ("good") or bello ("nice/beautiful"), so that the adjective always ended in /i/ or /ɛ/.

Vowel-initial words preceding nasal context		Vowel-initial words preceding vowel context	
Buon abito	*good dress*	Buoni abiti	*good dresses*
Buon aceto	*good vinegar*	Begli agnelli	*nice lambs*
Buon ago	*good needle*	Buoni aghi	*good needles*
Buon anello	*good ring*	Begli anelli	*nice rings*
Buon angelo	*good angel*	Begli angeli	*nice angels*
Buon arbitro	*good arbitrator*	Buone infermiere	*good nurses*
Buon armadio	*good cupboard*	Buoni armadi	*good cupboards*
Buon artista	*good artist*	Begli alberi	*nice trees*
Buon asciugamano	*good towel*	Begli asciugamani	*nice towels*
Buon attore	*good actor*	Begli attori	*nice actors*
Buon avocado	*good avocado*	Begli avocado	*nice avocados*
Buon opuscolo	*good brochure*	Begli elefanti	*nice elephants*

Buon elmo	*good helmet*	Begli elmi	*nice helmets*
Buon investigatore	*good detective*	Buoni investigatori	*good detectives*
Buon ombrello	*good umbrella*	Begli ombrelli	*nice umbrellas*
Buon orecchino	*good earring*	Begli orecchini	*nice earrings*
Buon orologio	*good watch*	Begli orologi	*nice watches*
Buon olio	*good oil*	Begli orsi	*nice bears*
Buon imbuto	*good funnel*	Belle isole	*nice islands*
Buon insegnante	*good teacher*	Begli uccelli	*nice birds*

Table III.B.3. Sentences used in eye-tracking Experiment 2a and 3a. Target words are in italics. All participants were presented all sentences from both list A and list B, but sentences in one of the two lists were presented in the substituted (Experiment 2a) or deleted (Experiment 3a) condition, and the respective other sentences in the correct version. Which list was presented in the correct condition was counterbalanced between participants. All words were presented with minimal context (underlined) in the rating tasks Experiment 2b and 3b in both conditions.

/h/-initial words in list A

Auf der Pizza mag sie am liebsten *Hackfleisch*.

On the pizza, she likes *minced meat* best.

Im Restaurant isst er gern *Hähnchen*.

At the restaurant, he likes to eat *chicken*.

Es wurde eiskalt nach dem vielen *Hagel*.

It became freezing after the heavy *hail*.

Ben überraschte seine Frau mit einer selbstgemachten *Halskette*.

Ben surprised his wife with a self-made *necklace*.

In seinem Rucksack hat Günter immer einen *Hammer*.

In his backpack Günter always has a *hammer*.

Zum Geburtstag bekam sie einen *Hamster*.

For her birthday, she was given a *hamster*.

Er gab ihr seinen *Handschuh*.

He gave her his *glove*.

Die Kinder spielen im *Haus*.
The children are playing in the *house*.

An der Wand hängt ein Bild mit einem *Heiligen*.
On the wall, there hangs a picture of a *saint*.

Gestern trug die Lehrerin einen *Helm*.
Yesterday the teacher wore a *helmet*.

Mama kauft immer die teuren *Hemden*.
Mom always buys the expensive *shirts*.

Urlaub hat sie erst im *Herbst*.
She has holiday only in *autumn*.

Katrin mag das Rezept mit den *Himbeeren*.
Katrin likes the recipe with the *raspberries*.

Er wirft den Ball in den *Himmel*.
He throws the ball into the *sky*.

Er ist genervt von der anhaltenden *Hitze*.
He is annoyed by the continuing *heat*.

Diese Tiere leben in den *Höhlen*.
These animals live in *caves*.

Das Gerüst machen sie aus teurem *Holz*.
They make the scaffold from expensive *wood*.

Zum Zentrum muss man nur vorbei an den *Hügeln*.
To get to the center you just have to pass the *hills*.

Die Camper übernachteten in *Hütten*.
The campers spent the night in *cottages*.

Das war ein Mann mit einem komischen *Hut*.
This was a man with a strange *hat*.

/h/-initial words in list B

Ihr Freund hat kaum *Haare*.
Her friend has hardly any *hair*.

Der Verein feierte damals viele Feste in der alten *Halle*.
At that time, the club celebrated many festivals in the old *hall*.

Sie hat starke Schmerzen im *Hals*.
She feels severe pain in her *throat*.

Sie verkaufen viel Bier bei den *Haltestellen*.
They sell a lot of beer at the *bus stops*.

Durch die Arbeit hat sie Schmerzen in den *Händen*.
Because of her work she suffers pain in her *hands*.

Jetzt liegt Tim auf dem *Handtuch*.
Now Tim is lying on the *towel*.

Es gibt immer technische Probleme mit dem *Handy*.
There are always technical problems with the *mobile phone*.

Max hört gern das Lied mit den *Harfen*.
Max likes to hear the songs with the *harps*.

Besonders süß sind die kleinen *Hasen*.
The little *hares* are particularly cute.

Der Lehrer erzählte uns von dem berühmten *Heer*.
The teacher told us about the famous *army*.

Das Wohnzimmer ist viel gemütlicher mit der neuen *Heizung*.
The living-room is much more comfortable with the new *heating*.

Jakob steht gerne am *Herd*.
Jakob likes to stand at the *stove*.

Toni mag das Spiel mit den *Herzen*.
Toni likes the game with the *hearts*.

Er feiert morgen *Hochzeit*.
He is celebrating his *wedding* tomorrow.

Abends isst er vor allem *Honig*.
In the evening, he mainly eats *honey*.

Man erkennt das Tier an den farbigen *Hörnern*.
You recognize the animal by the colored *horns*.

Sie trägt im Sommer am liebsten *Hosen*.
In summer, she prefers to wear *pants*.

Beim Tanzen hat sie Schmerzen in den *Hüften*.
When dancing, she feels pain in her *hips*.

Die Mädchen spielen mit den schönen *Hunden*.
The girls are playing with the beautiful *dogs*.

Das Kind spielt mit der neuen *Hupe*.

The child is playing with the new *horn*.

/ʔ/-initial words in list A

Tom trägt den schweren *Abfall* runter.
Tom carries down the heavy *garbage*.

Im Dschungel sieht er einen *Adler*.
In the jungle, he views an *eagle*.

Zum Reinigen nimmt Miriam *Alkohol*.
For cleaning, Miriam uses *alcohol*.

Es gibt oft Probleme mit den *Antennen*.
There are often problems with the *antennas*.

Am Nachmittag isst er einen *Apfel*.
In the afternoon, he eats an *apple*.

Zum Frühstück isst sie am liebsten *Aprikosen*.
For breakfast, she prefers to eat *apricots*.

Abends hat sie manchmal Schmerzen in den *Armen*.
In the evening, she sometimes suffers pain in her *arms*.

Heute trifft er zum ersten Mal einen echten *Arzt*.
Today he will meet a real *doctor* for the first time.

Er hatte nie den Berufswunsch, ein *Astronaut* zu sein.
He never wanted to become an *astronaut*.

Manchmal hat er Schmerzen im *Auge*.
Sometimes, he suffers pain in his *eye*.

Tobis Lieblingscocktail ist der mit den *Avokados*.
Tobi's favorite cocktail is the one with the *avocados*.

Sie wartet vor dem *Eingang*.
She is waiting in front of the *entrance*.

Abends mag er vor allem *Eis*.
In the evening, he mainly likes *ice cream*.

Das Kinderbuch handelt von einem *Engel*.
The children's book is about an *angel*.

Auf dem Bauernhof mag der Junge besonders die kleinen *Esel*.
On the farm, the little boy particularly likes the little *donkeys*.

Wir nehmen ein bisschen von dem guten *Öl*.
We take a some of the good *oil*.

Am Abend isst sie gern *Orangen*.
In the evening, she likes to eat *oranges*.

Der Koch braucht unbedingt einen neuen *Ordner*.
The cook really needs a new *folder*.

Im Sommer fährt sie gern *Ubahn*.
In summer, she likes to go by *metro*.

Er sieht gut aus in der neuen *Uniform*.
He looks good in the new *uniform*.

/ʔ/-initial words in list B

Im Dschungel sieht der Tourist einen großen *Affen*.
In the jungle, the tourist views a big *ape*.

Papa fährt bis zur großen *Ampel*.
Dad drives to the big *traffic lights*.

Zum Salat mag er gern *Ananas*.
He likes *pineapple* in his salad.

Vater trägt gern den warmen *Anzug*.
Father likes to wear the warm *suit*.

Der Junge reicht seinem Vater den *Aschenbecher*.
The boy passes his father the *ashtray*.

Sie hat jetzt einen neuen *Ausweis*.
Now she has got a new *identity card*.

Im Frühling machte Paul eine Tour mit seinem *Auto*.
In the spring, Paul goes on a tour in his *car*.

Vor dem Sport isst er gern *Eier*.
Before the sport, he likes to eat *eggs*.

Da ist ein Loch im *Eimer*.
There is a hole in the *bucket*.

Auf dem Bauernhof gefallen ihm die vielen *Enten*.
On the farm, he likes the many *ducks*.

In Deutschland isst Clara immer die guten *Erdbeeren*.
In Germany, Clara always eats the good *strawberries*.

Das Kinderbuch handelt von einem *Eskimo*.
The children's book is about an *Eskimo*.

Claudia bringt immer einen guten *Essig* mit.
Claudia always brings a good *vinegar*.

Am liebsten mag er die kleinen *Eulen*.
He likes the little *owls* best.

Morgens isst er am liebsten *Obst*.
In the morning, he prefers to eat *fruit*.

Der Schinken ist im *Ofen*.
The bacon is in the *oven*.

Sie hat manchmal Schmerzen im *Ohr*.
Sometimes, she feels pain in her *ear*.

Tanja ist gern *Oma*.
Tanja enjoys being a *granny*.

An der Ecke sieht sie den *Opa* stehen.
She views her *grandpa* standing at the corner.

Susanne liebt den großen *Ozean*.
Susanne loves the great *ocean*.

Appendix III.C: Pretest

Table III.C. Means and standard deviations (in brackets) of ratings given in the online pretest for all sentences used in the perception tasks. Values are shown for target and competitor words for the critical items /h/ and /ʔ/ in the two lists A and B, as well as for the filler sentences. In the pretest, sentences and words were presented in written form, and answers could be given on a five-point scale where 1 indicates that the word fits "very well" to the sentence and 5 means that the word "does not fit at all". Note that this coding is in line with the German marking system, in which the "1" is the best achievable mark and a "5" is a fail. Standard deviations indicate variability over the different sentences in the respective condition (see paragraph *Pretest* in Chapter 4, Experiment 2 for details).

List	Target word	Competitor word
/h/ A	2.16 (0.75)	1.47 (0.36)
/h/ B	2.24 (0.75)	1.54 (0.46)
/ʔ/ A	2.21 (0.81)	1.51 (0.44)
/ʔ/ B	2.17 (0.77)	1.40 (0.45)
fillers	1.20 (0.13)	2.62 (0.80)

BIBLIOGRAPHY

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in Speech motor programming. *Journal of Motor Behavior, 16*, 195-232. DOI: 10.1080/00222895.1984.10735318

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249-306. DOI: 10.1111/j.1467-9922.2009.00507.x

Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. *Fifth International Conference on Spoken Language Processing*, Paper 0429.
https://www.iscaspeech.org/archive/archive_papers/icslp_1998/ i98_0429.pdf

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419-439. DOI: 10.1006/jmla.1997.2558

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529-555. DOI: 10.1111/j.1467-1770.1992.tb01043.x

Aruffo, C., & Shore, D. I. (2012). Can you McGurk yourself? Self-face and self-voice in audiovisual speech. *Psychonomic Bulletin & Review, 19,* 66–72. DOI: 10.3758/s13423-011-0176-8

Asher, J. J., & García, R. (1969). The Optimal Age to Learn a Foreign Language. *The Modern Language Journal, 53*, 334-341. DOI: 10.1111/j.1540-4781.1969.tb04603.x

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*, 390-412. DOI: 10.1016/j.jml.2007.12.005

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied Linguistics in Language Teaching*, *44*, 231-250. DOI: 10.1515/IRAL.2006.010

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. DOI: 10.1016/j.jml.2012.11.001

Barry, W. J. (1979). Complex encoding in word-final voiced and voiceless stops. *Phonetica*, *36*, 361-372. DOI: 10.1159/000259973

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 191-206). Clevedon, UK: Multilingual Matters.

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1835-1842. DOI: 10.1037/xlm0000417

Bassetti, B., Sokolović-Perović, M., Mairano, P., & Cerni, T. (2018). Orthography-Induced Length Contrasts in the Second Language Phonological Systems of L2 Speakers of English: Evidence from Minimal Pairs. *Language and Speech*, 0023830918780141. DOI: 10.1177/0023830918780141

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1-51. DOI: 10.18637/jss.v067.i01

Becker, T. (2012). *Einführung in die Phonetik und Phonologie des Deutschen.* Darmstadt: Wissenschaftliche Buchgesellschaft.

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

Bent, T., Bradlow, A. R., & Smith, B. L. (2008). Production and perception of temporal patterns in native and non-native speech. *Phonetica, 65,* 131-147. DOI: 10.1159/000144077

Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*, 131-151. DOI: 10.1017/S0025100305002148

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*, 775. DOI: 10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-

speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345-360. DOI: 10.1037/0096-1523.14.3.345

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam, NL: John Benjamins. DOI: 10.1075/lllt.17.07bes

Bissiri, M. P., & Pfitzinger, H. R. (2009). Italian speakers learn lexical stress of German morphologically complex words. *Speech Communication, 51,* 933–947. DOI: 10.1016/j.specom.2009.03.001

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from http://www.praat.org/

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 279-304). Timonium MD: York Press.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*, 131-158. DOI: 10.1017/s0272263100010792

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

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299-2310. DOI: 10.1121/1.418276

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America, 117*, 3890-3901. DOI: 10.1121/1.1906060

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*, 1636-1644. DOI: 10.1121/1.3292996

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

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*, 22-34. DOI: 10.1016/j.system.2007.11.003

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*. 10.1027/1618-3169/a000123

Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, *22*, 1-26. DOI: 10.1017/s0272263100001017

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *The Journal of the Acoustical Society of America, 119*, 3085-3096. DOI: 10.1121/1.2188917

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*, 3647-3658. DOI: 10.1121/1.1815131

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, *11*, 1084-1089. DOI: 10.3758/BF03196741

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, *70*, 403-411. DOI: 10.3758/PP.70.3.403

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107. DOI: 10.1016/0010-0285(74)90005-X

Council of Europe (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Council of Europe. https://www.coe.int/en/web/common-european-framework-reference-languages, last viewed on 08/10/2018

Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, *36*, 115-128. DOI: 10.1017/s0142716414000459

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269-284. DOI: 10.1016/j.wocn.2005.06.002

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317-367. DOI: 10.1006/cogp.2001.0750

Darcy, I., Daidone, D., & Kojima, C. (2014). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*, 372-420. DOI: 10.1075/ml.8.3.06dar

de Groot, F., Huettig, F., & Olivers, C. N. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 180-196. DOI: 10.1037/xhp0000102

de Mareüil, P. B., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63*, 247-267. DOI: 10.1159/000097308

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research* (Vol. 42). John Benjamins Publishing Company: Amsterdam, Philadelphia.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*, 526-548. DOI: 10.1111/lang.12053

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, *27*, 47-55. DOI: 10.1017/S0025100300005417

Devue, C., & Brédart, S. (2011). The neural correlates of visual self-recognition. *Consciousness and Cognition, 20*, 40-51. DOI: 10.1016/j.concog.2010.09.007

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

Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. *Journal of Personality and Social Psychology*, *44*, 589-592. DOI: 10.1037/0022-3514.44.3.589

Draxler, C., & Jänsch, K. (2004). SpeechRecorder: A Universal Platform Independent Multi-Channel Audio Recording Software. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of Language Resources and Evaluation* (pp. 559-562). Lisbon, Portugal: Universidade Nova de Lisboa.

Eger, N. A., & Bohn, O. S. (2015). Picking up the cues to a new consonant contrast: Danish learners' production and perception of English word-final /s/-/z/. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of*

Phonetic Sciences. Glasgow, UK: the University of Glasgow. http://www.icphs2015.info/pdfs/Papers/ ICPHS0648.pdf

Eger, N. A., & Reinisch, E. (2019a). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 552-571. DOI: 10.1037/xlm0000599

Eger, N. A., & Reinisch, E. (2019b). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition, 41,* 179-200. DOI: 10.1017/S0272263117000377

Escudero, P., Benders, T., & Lipski, S.C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37,* 452-465. DOI: 10.1016/j.wocn.2009.07.006

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*, 345-360. DOI: 10.1016/j.wocn.2007.11.002

Eurostat, European Union, (2017). 60% of lower secondary level pupils studied more than one foreign language in 2015. Retrieved via http://ec.europa.eu/eurostat/documents/ 2995521/7879483/3-23022017-AP-EN.pdf/80715559-72ba-4c19-b341-7ddb42dd61a6, last viewed on 07/13/2018

Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, *139*, EL1-EL5. DOI: 10.1121/1.4939608

Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, *70*, 1-19. DOI: 10.1016/j.wocn.2018.05.003

Ferguson, S. H., Jongman, A., Sereno, J. A., & Keum, K. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology*, *21*, 153-162. DOI: 10.3766/jaaa.21.3.3

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R.* Los Angeles, CA: SAGE Publications Ltd.

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America, 76,* 692-707. DOI: 10.1121/1.391256

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. 233-277). Timonium MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. Schiller, & A. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–358). Berlin, Germany: Mouton de Gruyter. DOI: 10.1515/9783110895094.319

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

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish learners of English. *Journal of Phonetics*, *15*, 67-83.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America, 91,* 370-389. DOI: 10.1121/1.402780

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97,* 3125-3134. DOI: 10.1121/1.413041

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America, 92*, 128-143. DOI: 10.1121/1.404278

Gluszek, A., Newheiser, A. K., & Dovidio, J. F. (2011). Social psychological orientations and accent strength. *Journal of Language and Social Psychology, 30,* 28-45. DOI: 10.1177/0261927X10387100

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., & Bruneau, N. (2013). My voice or yours? An electrophysiological study. *Brain Topography, 26*, 72-82. DOI: 10.1007/s10548-012-0233-2

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*, 350-365. DOI: 10.1016/j.jcomdis.2006.06.013

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua–Spanish bilinguals. *Journal of Phonetics, 28,* 27-42. DOI: 10.1006/jpho.2000.0104

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception & Psychophysics, 74,* 613-629. DOI: 10.3758/s13414-017-1372-z

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29*, 163-180. DOI: 10.1177/0265532211421161

Hattori, K., & Iverson, P. (2008). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*, 469-479. DOI: 10.1121/1.3021295

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English /r/-/l/ perception and production by adult Japanese speakers. In *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*. Tokyo: Waseda University.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*, 5-33. DOI: 10.1177/0267658307082980

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*, 367–381. DOI: 10.1177/0023830910371460

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics, 36*, 664-679. DOI: 10.1016/j.wocn.2008.04.002

Herd, W., Jongman, A., & Sereno, J. A. (2013). Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish. *The Journal of the Acoustical Society of America, 133*, 4247-4255. DOI: 10.1121/1.4802902

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131-138. DOI: 10.1016/S1364-6613(00)01463-7

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*, 3099-3111. DOI: 10.1121/1.411872

Hobel, B., Moosmüller, S., & Kaseß, C. (2016). Pronunciation norms and pronunciation habits of orthographic <ä, äh> in Standard Austrian German. *Phonetician, 113,* 24-48.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, *45*, 295-310. DOI: 10.1044/1092-4388(2002/023)

Howell, P., & Dworzynski, K. (2001). Strength of German accent under altered auditory feedback. *Perception & Psychophysics*, *63*, 501-513. DOI: 10.3758/BF03194416

Huan, B. H., & Jun, S.-A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech, 54*, 387-414. DOI: 10.1177/0023830911402599

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*, 460-482. DOI: 10.1016/j.jml.2007.02.001

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151-171. DOI: 10.1016/j.actpsy.2010.11.003

Hughes, A., Trudgill, P., & Watt, D. (2012). *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles*. 5th edition. London: Hodder Education.

Imai, S., Flege, J. E., & Walley, A. (2003). The recognition of accented and unaccented English words by native speakers of Spanish and English. *The Journal of the Acoustical Society of America, 113*, 2255. DOI: 10.1121/1.4780439

Ingram, J. C., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343-370. DOI: 10.1006/jpho.1997.0048

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. DOI: 10.1017/S1366728911000447

Ionta, S., Gassert, R., & Blanke, O. (2011). Multi-Sensory and Sensorimotor Foundation of Bodily Self-Consciousness – An Interdisciplinary Approach. *Frontiers in Psychology, 2*, 113-120. DOI: 10.3389/fpsyg.2011.00383

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57. DOI: 10.1016/S0010-0277(02)00198-1

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446. DOI: 10.1016/j.jml.2007.11.007

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*, 206-219.

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, *3*, 218-223. DOI: 10.1093/scan/nsn014

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246. DOI: 10.3389/fpsyg.2014.01246

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

Kassaian, Z. (2011). Age and gender effect in phonetic perception and production. *Journal of Language Teaching Research*, *2*, 370–376. DOI: 10.4304/jltr.2.2.370-376

Kisler, T., Reichel, U. D., & Schiel, F. (2017): Multilingual processing of speech via web services. *Computer Speech & Language, 45*, 326–347.

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception in incomplete neutralization of final devoicing in German. *Journal of Phonetics, 38*, 185-195. DOI: 10.1016/j.wocn.2009.10.001

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*, 148. DOI: 10.1037/a0038695

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). The Relationship between the Perception and Production of English Nasal Codas by Brazilian Learners of English. In *Proceedings of Interspeech* (pp. 2297-2300). Antwerp, Belgium.

Kohler, K. J. (1994). Glottal stops and glottalization in German. *Phonetica*, *51*, 38-51. DOI: 10.1159/000261957

Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The pros and cons of dubbing and subtitling. *European Journal of Communication*, *17*, 325-354. DOI: 10.1177/0267323102017003694

Krämer, M. (2009). *The phonology of Italian.* Oxford: University Press.

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, *25*, 2179-2188. DOI: 10.1162/jocn_a_00463

Kubozono, H. (2002). Prosodic Structure of Loanwords in Japanese: Syllable Structure, Accent and Morphology. *Journal of the Phonetic Society of Japan*, *6*, 79-97. DOI: 10.24467/onseikenkyu.6.1_79

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 979-1000. DOI: 10.1098/rstb.2007.2154

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, Cambridge: Blackwell Publishers.

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, *34*, 10339-10346. DOI: 10.1523/JNEUROSCI.0108-14.2014

Lenneberg, E. (1967). *Biological Foundations of Language.* New York: Wiley.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75. DOI: 10.1017/S0140525X99221772

Levy, E. S., & Law II, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*, 1290-1305. DOI: 10.1121/1.3466879

Li, G., & Mok, P. P. K. (2015). Interlanguage Speech Intelligibility Benefit for Mandarin: Is it from shared phonological knowledge or exposure to accented speech. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0509.pdf

Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1040-1056. DOI: http://dx.doi.org/10.1037/xhp0000383

Llompart, M., & Reinisch, E. (in press_a). The robustness of lexical representations in a second language relates to phonetic flexibility for difficult sound contrasts. *Bilingualism: Language and Cognition.* DOI:10.1017/S1366728918000925

Llompart, M., & Reinisch, E. (in press_b). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*. DOI: 10.1177/0023830918803978

Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, *26*, 381-400. DOI: 10.1006/jpho.1998.0081

Magno-Caldognetto, E. M., Zmarich, C., & Ferrero, F. (1997). A comparative acoustic study of spontaneous and read Italian speech. In *Fifth European Conference on Speech Communication and Technology*, 779-782.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*, 173-190. DOI: 10.2307/3588329

Malécot, A. (1975). The glottal stop in French. *Phonetica*, *31*, 51-63. DOI: 10.1159/000259649

Marotta, G. (2008). Lenition in Tuscan Italian (Gorgia Toscana). In J. Brandão de Carvalho, T. Scheer, & P. Ségéral (Eds.), *Lenition and fortition* (pp. 235-271). Berlin: Mouton de Gruyter.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*, 154-178. DOI: 10.1002/dev.21177

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*, 229-258. DOI: 10.1006/jpho.2002.0174

McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*, 317-339. DOI: 10.1080/15475441.2011.641887

Mitsuya, T., Samson, F., Ménard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, *133*, 2993-3003. DOI: 10.1121/1.4795786

Mitterer, H. (2018). Not all geminates are created equal: Evidence from Maltese glottal consonsants. *Journal of Phonetics, 66*, 28-44. DOI: 10.1016/j.wocn.2017.09.003

Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language, 85,* 116-134. DOI: 10.1016/j.jml.2015.08.005

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323-331. DOI: 10.1016/0010-0277(79)90020-9

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61-64. DOI: 10.20982/tqmp.04.2.p061

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *14*, 477-492. DOI: 10.1016/0022-0965(72)90066-5

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Multicultural Development, 28,* 502-518. DOI: 10.2167/jmmd514.0

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49,* 285-310. DOI: 10.1111/0023-8333.49.s1.8

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111-131. DOI: 10.1017/S0272263106060049

Munro, M., & Mann, V. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics, 26*, 311-341. DOI: 10.1017/S0142716405050198

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). Feedback-driven corrective movements in speech in the absence of altered feedback. *The Journal of the Acoustical Society of America*, *134*, 4167-4167. DOI: 10.1121/1.4831274

Ohk, B. (2006). *Dizzjunarju Ġermaniż – Malti: Wörterbuch Deutsch – Maltesisch.* Marsa, Malta: Grima Print.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*, B9-B17. DOI: 10.1016/S0010-0277(97)00030-9

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, *12*, 445-449. DOI: 10.1111/1467-9280.00383

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11*, 73-89.

Pattamadilok, C., Morais, J., Colin, C., & Kolinsky, R. (2014). Unattentive speech processing is influenced by orthographic knowledge: Evidence from mismatch negativity. *Brain and Language*, *137*, 103-111. DOI: 10.1016/j.bandl.2014.08.005

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, *58*, 254-275. DOI: 10.1159/000046178

Peabody, M., & Seneff, S. (2006). Towards Automatic Tone Correction in Non-native Mandarin. In: Q. Huo, B. Ma, E.-S. Chng, & H. Li (Eds.), *Chinese Spoken Language Processing. ISCSLP 2006. Lecture Notes in Computer Science, 4274* (pp. 602-613). Springer: Berlin, Heidelberg.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13. DOI: 10.1016/j.jneumeth.2006.11.017

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of Interspeech* (pp. 161-164). Florence, Italy.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America, 116*, 2338-2344. DOI: 10.1121/1.1787424

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. DOI: 10.1121/1.3613698

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29,* 191-215. DOI: 10.1006/jpho.2001.0134

Platek, S. M., Burch, R. L., & Gallup, G. G. (2001). Sex differences in olfactory self-recognition. *Physiology & Behavior, 73*, 635-640. DOI: 10.1016/s0031-9384(01)00539-x

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, *32*, 141-152. DOI: 10.3758/bf03204273

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*, 97-132. DOI: 10.1016/s0010-0277(00)00090-1

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413-425. DOI: 10.1016/j.jml.2008.02.002

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718-724. DOI: 10.1038/nn.2331

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Reinisch, E. (2016a). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics, 78*, 1203-1217. DOI: 10.3758/s13414-016-1067-x

Reinisch, E. (2016b). Speaker-specific processing and local context information: the case of speaking rate. *Applied Psycholinguistics*, 37, 1397-1415. DOI: http://dx.doi.org/10.1017/S0142716415000612

Reinisch, E., & Weber, A. (2012). Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America, 132*, 1165-1176. DOI: 10.1121/1.4730884

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

Rindal, U. (2010). Constructing identity with L2: Pronunciation and attitudes among Norwegian learners of English. *Journal of Sociolinguistics*, *14*, 240-261. DOI: 10.1111/j.1467-9841.2010.00442.x

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics, 43*, 11-25. DOI: 10.1016/j.wocn.2014.01.002

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition, 68*, 204-214. DOI: 10.1016/j.bandc.2008.04.007

Rose, M. (2010). Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices* (pp. 181-196). Somerville, MA: Cascadilla.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62,* 595-633. DOI: 10.1111/j.1467-9922.2011.00639.x

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. DOI: 10.1017/S0142716417000418

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*, 1207-1218. DOI: 10.3758/APP.71.6.1207

Scarpace, D. (2014). The Acquisition of the Tap/Trill Contrast Within and Across Words in Spanish. In W. Cardoso, L. Kurtz dos Santos Buss, & P. Kielstra (Eds.), *Proceedings of the International Symposium on the Acquisition of Second Language Speech*, Volume 5 (pp. 580-596). Montreal: Concordia Working Papers in Applied Linguistics. http://doe.concordia.ca/copal/documents/37_Scarpace_Vol5.pdf

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183-204. DOI: 10.1016/j.wocn.2015.07.003

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*, 355-367. DOI: 10.3758/s13414-015-0987-1

Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing, 31,* 367-388. DOI: 10.1177/0265532214526175

Schuerman, W. L. (2017). *Sensorimotor Experience in Speech Perception.* Doctoral dissertation. Radboud University Nijmegen.

Schuerman, W. L., Meyer, A., & McQueen, J. M. (2015). Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. *PLoS ONE*, 10: e0129731. DOI: 10.1371/journal.pone.0129731

Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning*, *62*, 131-147. DOI: 10.1111/j.1467-9922.2012.00709.x

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240-255. DOI: 10.1016/j.jml.2004.11.001

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. doi: 10.1017/S0142716400001417

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*, 1103-1113. DOI: 10.1121/1.3058638

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research, 41*, 941-950. DOI: 10.1044/jslhr.4104.941

Shuster, L. I., & Durrant, J. D. (2003). Toward a better understanding of the perception of self-produced speech. *Journal of Communication Disorders*, *36*, 1-11. DOI: 10.1016/s0021-9924(02)00132-6

Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America, 125*, 3306-3316. DOI: 10.1121/1.3101452

Simonchyk, A., & Darcy, I. (2018). The effect of orthography on the lexical encoding of palatalized consonants in L2 Russian. *Language and Speech, 61,* 522-546. DOI: 10.1177/0023830918761490

Smith, B. L., & Hayes-Harb, R. (2011). Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of Phonetics*, *39*, 115-120. DOI: 10.1016/j.wocn.2010.11.005

Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, *37*, 257-275. DOI: 10.1016/j.wocn.2009.03.001

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281-284. DOI: 10.1111/1467-9280.00151

SR Research Experiment Builder 1.10.1630 [Computer software]. (2011). Mississauga, Ontario, Canada: SR Research Ltd.

Stevens, M., Hajek, J., & Absalom, M. (2002). Raddoppiamento sintattico and glottalization phenomena in Italian: a first phonetic excursus. In C. Bow (Ed.), *Proceedings of the 9th Australian International Conference on Speech Science & Technology* (pp. 154-159). Melbourne: Australian Speech Science & Technology Association Inc.

Stibbart, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America, 120*, 433. DOI: 10.1121/1.2203595

Strömbergsson, S., Wengelin, Å., & House, D. (2014). Children's perception of their synthetically corrected speech production. *Clinical Linguistics & Phonetics*, *28*, 373-395. DOI: 10.3109/02699206.2013.868928

Stuart-Smith, J. (2007). The influence of the media. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge Companion to Sociolinguistics* (pp. 140-148). London, New York: Routledge.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 41,* 177-204. DOI: 10.1111/j.1467-1770.1991.tb00683.x

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: a computer-mediated approach. *Language Learning, 62,* 1231-1258. DOI: 10.1111/j.1467-9922.2012.00724.x

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*, 952-981. DOI: 10.1080/01690960903498424

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Prag: Cercle Linguistique.

Trudgill, P. (2014). Diffusion, drift, and the irrelevance of media influence. *Journal of Sociolinguistics*, *18*, 213-222. DOI: 10.1111/josl.12070

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*, 263-290. DOI: 10.1016/j.wocn.2004.10.002

Underbakke, M. E. (1993). Hearing the difference: Improving Japanese students' pronunciation of a second language through listening. *Language Quarterly, 31*, 67–89.

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*, 1000. DOI: 10.3389/fpsyg.2015.01000

van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*, 181-198. DOI: 10.3758/BF03197716

van Santen, J., & D'Imperio, M. (1999). Positional effects on stressed vowel duration in Standard Italian. In J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A.C. Bailey (Eds.), *Proceedings of the 14th International Congress of the Phonetic Sciences* (pp. 1757-1760). San Francisco.
https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0241.pdf

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35,* 103-113. DOI: 10.1016/S0167-6393(00)00098-4

van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America, 111,* 1906-1916. DOI: 10.1121/1.1456928

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica, 64,* 122-144. DOI: 10.1159/000107913

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*, 479-491. DOI: 10.1016/j.wocn.2010.12.004

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1-25. DOI: 10.1016/S0749-596X(03)00105-0

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34-51. DOI: 10.1016/j.wocn.2014.05.002

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49-63. DOI: 10.1016/S0163-6383(84)80022-3

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

White, E. J., Titone, D., Genesee, F., & Steinhauer, K. (2017). Phonological processing in late second language learners: The effects of proficiency and task. *Bilingualism: Language and Cognition*, *20*, 162-183. DOI: 10.1017/S1366728915000620

Wiese, R. (1996). *The phonology of German.* Oxford: Clarendon Press.

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing, 30,* 231-252. DOI:10.1177/0265532212456968

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*, 537-556. DOI: 10.3758/s13414-012-0404-y

Wong, J. W. S. (2013). The effects of perceptual and or productive training on the perception and production of English vowels /I/ and /i:/ by Cantonese ESL learners. In *14th Annual Conference of the International Speech Communication Association* (pp. 2113–2117).
https://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_2113.pdf

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 34-57). Cambridge: University Press.

Xie, X., & Fowler, C. A. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics, 41*, 369-378. DOI: 10.1016/j.wocn.2013.06.003

Xu, M., Homae, F., Hashimoto, R. I., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology, 4*. DOI: 10.3389/fpsyg.2013.00735

Yamada, R.A. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech Perception and Lingustic Experience: Issues in Cross-language Research* (pp. 305-320). Timonium MD: York Press.

Ziegler, W. (2010). Dysarthrie. In G. Blanken & R. de Bleser (Eds.), *Mentale Sprachverarbeitung 6* (pp. 257–281). Mainz, Aachen: HochschulVerlag.

Zimmerer, F., & Trouvain, J. (2015). “Das Haus“ or “das Aus“? – How French learners produce word-initial /h/ in German. In A. Leemann, M.-J. Kolly, S. Schmid, & V. Dellwo (Eds.), *Trends in Phonetics and Phonology. Studies from German-speaking Europe* (pp. 303-316). Bern: Peter Lang.