

The role of acoustic cues and listener proficiency in the perception of accent in non-native sounds

Nikola Anna Eger & Eva Reinisch

Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich
eger@phonetik.uni-muenchen.de, evarei@phonetik.uni-muenchen.de

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 non-native 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.

Keywords: Foreign accent; Speech perception; Acoustic cues; Listener proficiency

Running head: Acoustic cues and proficiency in accent perception

Accepted for publication at *Studies in Second Language Acquisition*

©Cambridge University Press

Published online: 18 December 2017

doi: 10.1017/S0272263117000377

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 non-native 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 non-native 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 non-native 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; Inghvalson, 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 non-native 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 non-native 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 / ϵ / – / 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 vowel¹, which is acoustically and articulatorily close to English / ϵ /. 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 / ϵ / and as a consequence often also produce it as / ϵ -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-finally². 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 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.

METHOD

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".

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 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 sentences³. Twenty-four female⁴ 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 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 B.1 to B.3 in the Appendix.

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 A). The selection proceeded as follows: first, words with other difficult sounds than the critical contrast were excluded (e.g., words with the contrast /ε/-/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.

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.

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 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 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 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 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).

(Insert Figure 1 about here)

RESULTS

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 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).

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 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 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 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 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.

(Insert Table 1 about here)

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 non-native 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 non-native 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 non-native 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 non-native 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; Llompарт & 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.

REFERENCES

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

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

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

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/llt.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., & 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

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

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

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

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

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

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

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

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

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., & 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

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

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

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

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.

Imai, S., Flege, J., & 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

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

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

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 38, 185-196. DOI: 10.1016/j.wocn.2009.10.001

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

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

Llompert, 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>

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

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

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

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

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

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

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

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

Röttger, 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

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

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

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

Sidas, 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

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

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

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

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

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 co-determine speed of perceptual adaptation. *Attention, Perception & Psychophysics, 75*, 537-556. DOI: 10.3758/s13414-012-0404-y

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

NOTES

¹Orthographically 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).

²Note that this neutralization in German has been shown to be incomplete (e.g., Röttger, Winter, Grawunder, Kirby, & Grice, 2014). However, German listeners do not use this information directly in perception (Kleber, John, & Harrington, 2010).

³The 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.

⁴Only female speakers were recruited to focus listeners' attention on the pronunciation of the critical words/sounds rather than differences in voice (quality).

Tables

Table 1: Results of the mixed-effects model fitted with Sound Type, Material, listener Proficiency, and their interactions for the German learners.

Fixed effect	<i>b</i>	SE	df	<i>t</i>	<i>p</i>
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

Figure Captions

Figure 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).

Figures

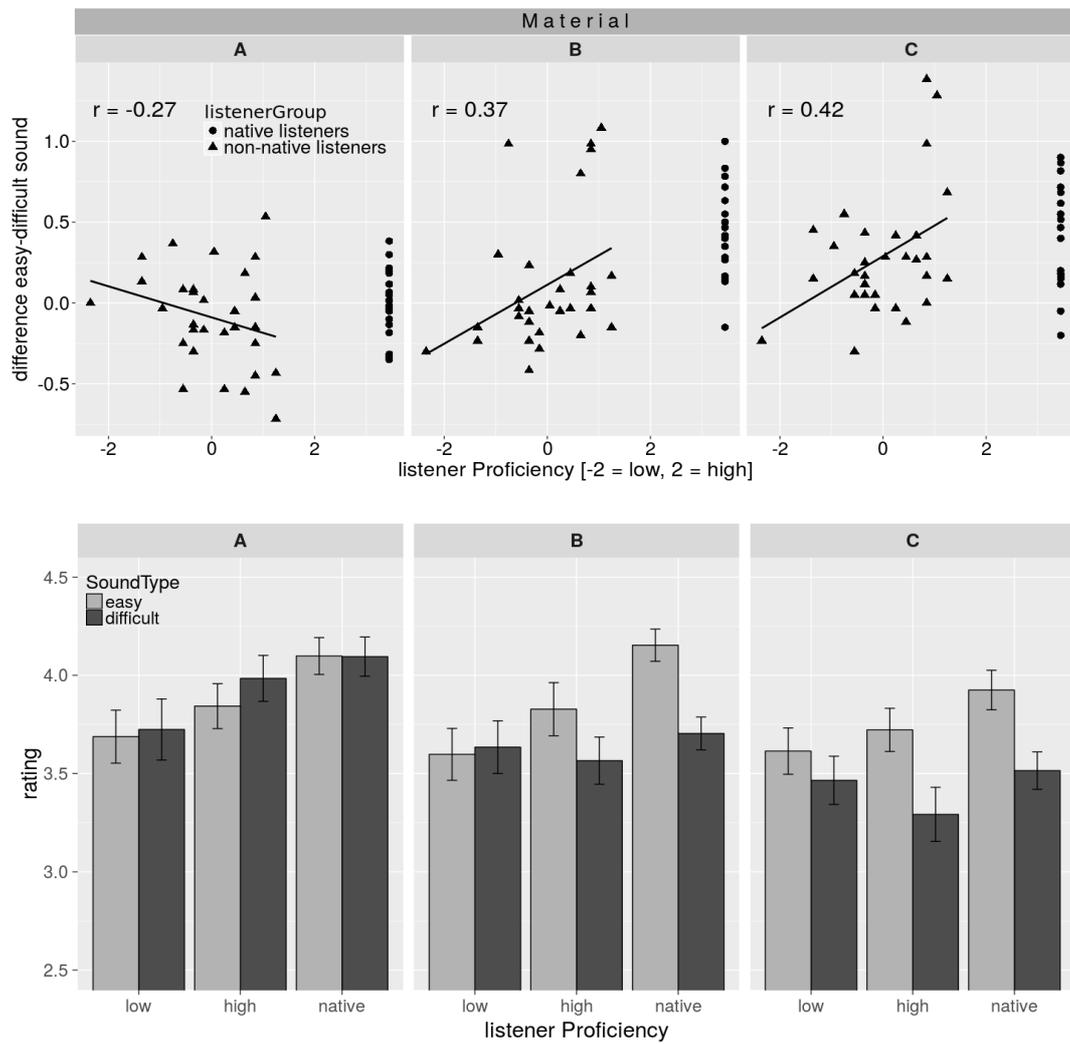


Figure 1

APPENDIX

Appendix A

Table 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
<i>bed – bad</i>	<i>face – phase</i>	<i>back – bag</i>	get	state
<i>dead – Dad</i>	<i>proof – prove</i>	<i>bat – bad</i>	honest	strong
<i>head – had</i>		<i>bet – bed</i>	king	time
<i>letter – latter</i>		<i>bright – bride</i>	kiss	worse
<i>merry – marry</i>		<i>brought – broad</i>	nine	worth
<i>send – sand</i>		<i>heart – hard</i>		
		<i>height – hide</i>		
		<i>white – wide</i>		

Appendix B

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).

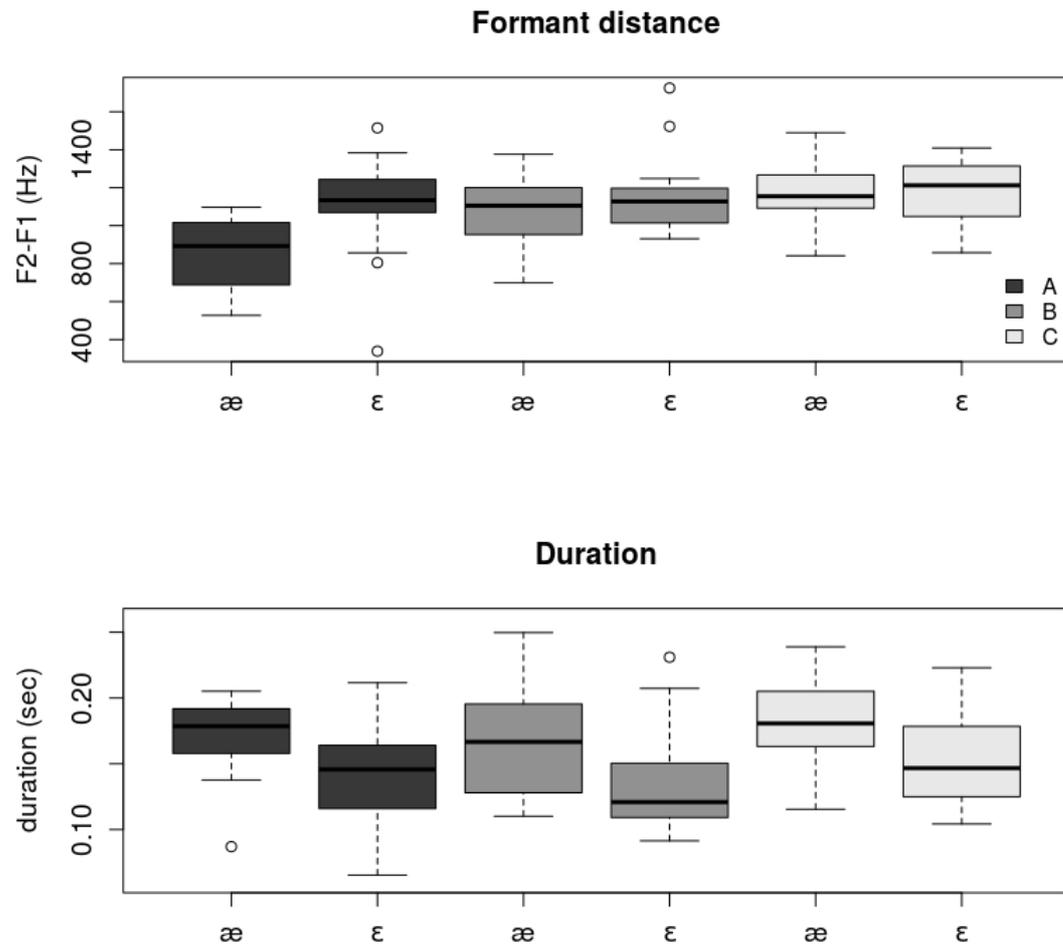


Figure B.1 Upper Panel: Formant values measured as the difference between F2 and F1 in Hz during a stable segment in the vowel for words with either /æ/ or /ɛ/ for the German learners grouped into three groups of 4 (dark gray = group A, mid gray = group B, light gray = group C); Lower Panel: Duration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

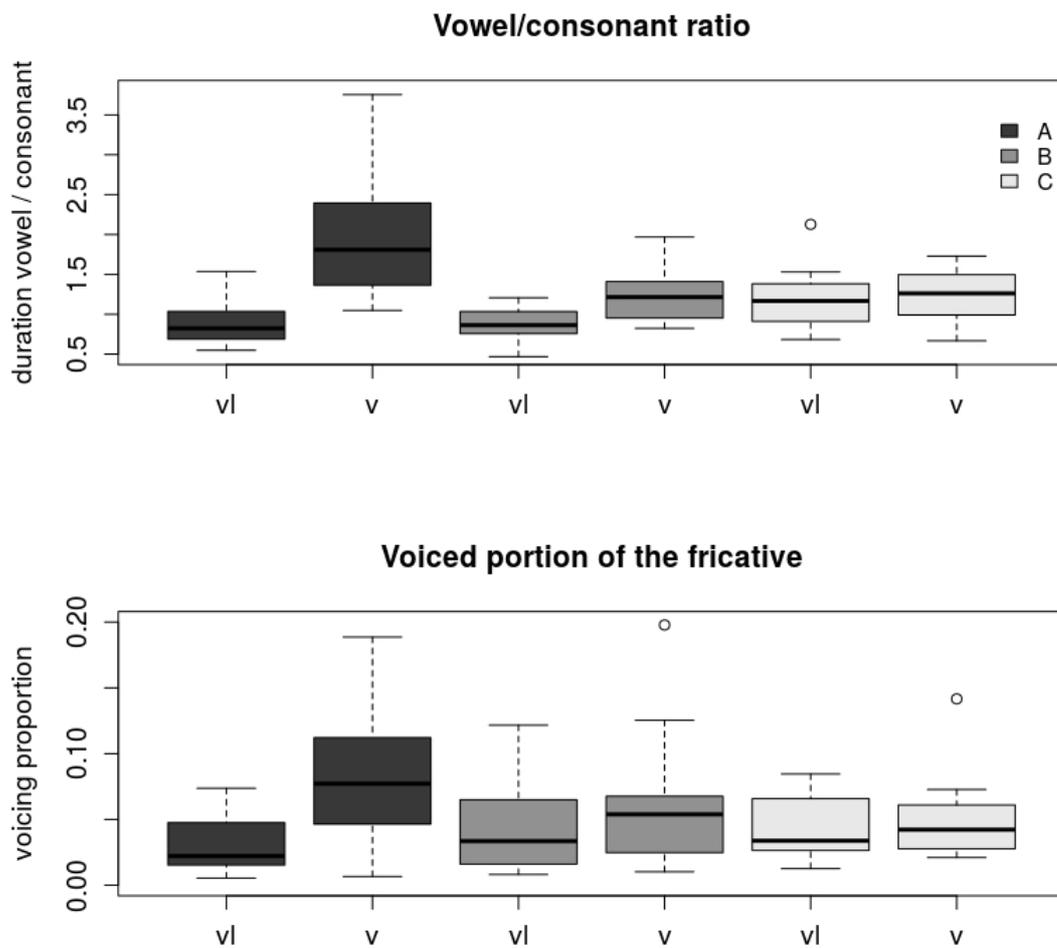


Figure B.2 Upper panel: Vowel/consonant ratios measured as the duration of the vowel divided by the duration of the consonant in words ending in voiced (v) or voiceless (vl) fricatives for the German learners grouped into three groups of 4 (dark gray = group A, mid gray = group B, light gray = group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

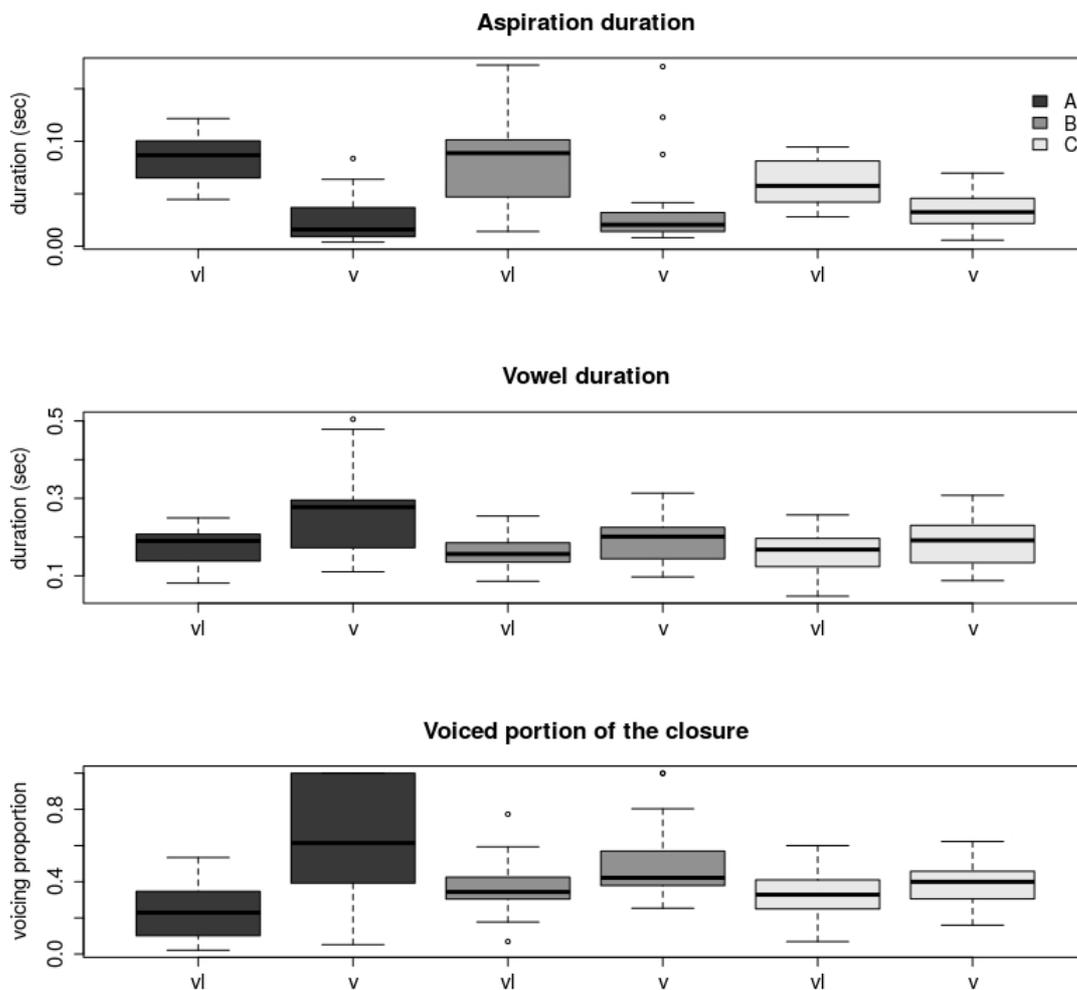


Figure B.3 Top panel: Aspiration duration for words ending in either voiced (v) or voiceless (vl) stops for the German learners grouped into three groups of 4 (dark gray = group A, mid gray = group B, light gray = group C); Mid panel: Duration of the preceding vowel; Bottom panel: Voiced portion of the 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.