



Research Article

## L2 immersion causes non-native-like L1 pronunciation in German attriters

Christopher Bergmann <sup>a,\*</sup>, Amber Nota <sup>a</sup>, Simone A. Sprenger <sup>a</sup>, Monika S. Schmid <sup>b</sup>

<sup>a</sup> Department of English Linguistics, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands

<sup>b</sup> Centre for Research in Language Development throughout the Lifespan, Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom



ARTICLE INFO

Article history:

Received 4 February 2015

Received in revised form

17 June 2016

Accepted 8 July 2016

Available online 29 July 2016

Keywords:

Language attrition

Bilingualism

Segmental articulation

Speech Learning Model

German

English

ABSTRACT

According to Flege's *Speech Learning Model*, the speech sounds of a bilingual's languages are contained in one common phonological space. This predicts bidirectional influence on the articulation of these speech sounds. We investigated the influence of a late-learned second language (L2) on the first language (L1) in a group of German L1 attriters in Anglophone North America (i.e., long-term emigrants in L2 immersion). These speakers were compared to a control group of monolingual German L1 speakers in two analyses: First, L1 speech samples of both groups were rated for native-likeness. Attriters sounded less native-like to raters, with 40% of the attriters rated below the monolingual range. Native-likeness was negatively associated with length of residence abroad and positively associated with L1 use. Second, formant analyses on four speech sounds of German—/a:/, /ɛ/, /ɔ/ and /ɪ/—were conducted for attriters and controls. For these analyses, two attriter subgroups were formed: one with speakers who sounded native-like to raters and one with speakers who did not. It was hypothesised that the formants in both groups would shift in the direction of similar L2 speech sounds and that the shift would be stronger in non-native-like attriters. The first hypothesis was partly confirmed: At least one attriter group differed from the control group on one formant of /a:/ and /ɪ/. These differences were consistent with predictions based on the L2. The second hypothesis was not confirmed: There was no evidence that the formants of the non-native-like attriters deviated more strongly from the monolingual baseline than those of the native-like attriters. Additionally, the formant values and the ratings were found to be only weakly associated, suggesting a different source of the perceptibly non-native-like pronunciation in some attriters.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

In their second language (L2), many speakers have an accent that is shaped by their first language (L1) and is therefore perceived as non-native-like. It is less well attested if and under what conditions prolonged exposure to an L2 influences the way one pronounces L1 speech sounds. Flege's (1995) *Speech Learning Model* assumes that the speech sounds of a bilingual's languages are represented in one common phonological space. In this space, bidirectional influence between the sounds of both languages is likely to occur, predicting articulatory changes in the L1 in a similar fashion as in the L2. This may result in the assimilation of the articulation of individual speech sounds towards the L2 settings. Questions arise regarding how the articulation of individual speech sounds, as represented in the phonological space, relates to the perceived global native-likeness of a speaker's pronunciation.

In this study, we focused on the spontaneous speech of a group of late bilinguals, namely L1 German speakers who, as adults, emigrated to a North-American L2 English environment. Potential changes in their L1 pronunciation were investigated from two angles: We first tried to establish if immersion in an L2 environment leads to global changes in pronunciation. This was assessed by means of a rating study in which native raters judged speech samples by these bilingual speakers and predominantly monolingual controls. The results of such a rating study, however, cannot tell us why some speakers sound more native-like than others: Do they pronounce individual speech sounds differently? To gain more insight into this, we then compared one group of monolinguals and two

\* Corresponding author.

E-mail addresses: [c.bergmann@rug.nl](mailto:c.bergmann@rug.nl) (C. Bergmann), [a.g.r.nota@rug.nl](mailto:a.g.r.nota@rug.nl) (A. Nota), [s.a.sprenger@rug.nl](mailto:s.a.sprenger@rug.nl) (S.A. Sprenger), [mschmid@essex.ac.uk](mailto:mschmid@essex.ac.uk) (M.S. Schmid).

groups of bilinguals—one rated as sounding native-like, the other as non-native-like—on the segmental articulation of selected phonemes. If changes in segmental articulation are the main source of sounding non-native-like, the groups should differ from one another in how far they deviate from monolinguals in their pronunciation of the speech sounds we analysed. Also, it should be possible to establish a relationship between the native-likeness ratings and the articulatory deviance in individual speakers.

Furthermore, the native-likeness of a speaker's pronunciation—both globally and on the level of individual speech sounds—is likely to be related to individual characteristics of this speaker, such as general L1 proficiency, frequency of L1 use and length of residence in the L2 environment. Therefore, we also determined if any of these factors was associated with the results of our analyses.

### 1.1. Speech Learning Model

The *Speech Learning Model* (SLM; [Flege, Schirru, & MacKay, 2003](#); [Flege, 1995, 2002](#)) is primarily a model of L2 phonetic acquisition, formulated to account for the fact that individuals do not produce speech sounds in their L2 in a native-like fashion. The model is interesting for the investigation of L1 speech of bilinguals because it also postulates that the phonetic categories of both languages of a bilingual exist in a common phonological space, where they mutually influence each other. A common phonological space for all languages a bilingual speaks and an influence of those languages on one another are also assumed by other models of bilingual sound production, such as the *Perceptual Assimilation Model* ([Best, 1995](#)) and the *Native Language Magnet Model* ([Kuhl, 1993, 2000](#)).

In the context of L2 influence on L1 speech, two of the hypotheses derived from the SLM are most relevant: First, it is hypothesised that during L2 acquisition, no new categories are formed for L2 speech sounds that are perceived as similar to an L1 sound. Rather, the L2 speech sound is understood as a variant of an L1 sound; the relationship between the speech sounds is formed on an allophonic (rather than phonemic) level. This equivalence classification is based on individual perception, linking phones that sound most similar to a speaker. Speech sounds that have been linked this way are called 'diaphones' ([Weinreich, 1957](#)). Second, the SLM predicts 'category assimilation' in diaphones: This means that the mapping of L1 and L2 speech sounds onto one another leads to these sounds coming to resemble each other in production. It is proposed that one single long-term memory representation for diaphones evolves, yielding a 'merged' category for the L1 and L2 speech sound. The SLM not only predicts that L2 sounds are influenced by L1 diaphones, but also that L1 sounds are realised differently when they have been linked to L2 sounds.

### 1.2. What is language attrition?

Language attrition refers to changes (usually a decline) in an individual's abilities in a language, induced by decreased use of and input in this language. L1 attrition is commonly experienced in the context of migration when people move to an environment in which their mother tongue is not (widely) spoken ([Schmid, 2004](#)). L1 attritors often become very proficient in their L2—the language they use every day—and in some cases, language dominance reversal in favour of the L2 can take place ([Mägiste, 1979](#); [Opitz, 2010](#)).

The individuals we are interested in are post-puberty L1 attritors (i.e., late bilinguals). When investigating the L2 of late learners, it can be difficult to trace differences from L1 speakers to discrete factors: Potential maturational constraints (see [Newport, Bavelier, and Neville \(2011\)](#) for a critical overview), socio-psychological aspects ([Flege et al., 2006](#)) and language competition between the L1 and the L2 ([Bergmann, Sprenger, & Schmid, 2015](#); [Dijkstra, 2005](#); [Kroll, Bobb, Misra, & Guo, 2008](#)) have been discussed as causes of performance differences between these bilingual and monolingual speakers. By contrast, late L1 attritors have acquired their L1 during infancy and childhood, only emigrating at an adult age after having reached full monolingual native proficiency. They have started from the same level of L1 mastery as other monolinguals and maturational factors do not apply to them. The only factor distinguishing them from monolingual speakers who have never emigrated from their home country is the mental presence of a permanently active L2 ([Kroll, Bobb, & Wodniecka, 2006](#)), possibly interfering with their L1. In these speakers, it is therefore easier than in L2 learners to isolate the effects of bilingualism from other factors constraining language development ([Schmid, 2014](#)).

L1 attrition in late emigrants has been found in a variety of domains: both in production ([Schmid, 2002](#)) and in perception ([Major, 2010](#)) as well as across a wide gamut of linguistic levels from lexical ([Köpke, 2002](#); [Pavlenko, 2000](#); [Schmid & Jarvis, 2014](#)) and morphological ([Keijzer, 2010](#); [Ribbert & Kuiken, 2010](#); [Schmitt, 2010](#)) to syntactic ([Gürel, 2004, 2008](#); [Tsimpli, Sorace, Heycock, & Filaci, 2004](#)) and pragmatic phenomena ([Brown & Gullberg, 2008, 2011](#)). It is conspicuous that in many studies, a subgroup of emigrants is found to remain native-like and even speakers who do experience attrition perform well, often better than highly proficient L2 speakers.

In the following, we review the existing evidence of attrition in the phonetic domain, focusing on global accent ratings and acoustical investigations of segmental phenomena. Most of these studies, similarly to the present one, investigate language attrition as a surface phenomenon, that is, as linguistic behaviour recorded at a given moment without any implications about the permanence of the findings. Attrition must not be understood as permanent 'loss', but rather as a change in L1 performance that is attributed to a language dominance reversal in favour of the L2 or simply to bilingualism as such ([Paradis, 2004, 2007](#)).

### 1.3. Literature review: phonetic attrition

Within the literature on phonetic attrition, only one publication that we are aware of reports the results of both native-likeness ratings and acoustical measurements on the speech of the same individual. It is a case study of a speaker of L1 Brazilian Portuguese and L2 American English (Sancier & Fowler, 1997). In her L1, the speaker sounded more accented to native raters after prolonged stays in the L2 environment. Also, the voice onset time (see below) of her L1 voiceless stops was longer (i.e., more English-like) after US stays. In this case, we probably see an example of temporary change due to language interference, rather than actual loss.

To learn more about potentially L2-induced changes in global pronunciation and in the articulation of individual speech sounds, we turn to research that has only investigated either the one or the other.

#### 1.3.1. Native-likeness ratings

Rating studies in which monolingual speakers judge the pronunciation of other speakers offer the most insight into the global, perceptible phonetic differences between monolingual and bilingual speakers. The number of such studies is fairly limited, though. To the best of our knowledge, all currently existing publications in this domain focus on the L1 speech of native speakers of German.

A group of L1 German bilinguals in L2 English and L2 Dutch immersion contexts, for example, was found to be rated as sounding less native-like in spontaneous speech samples than a control group of German monolinguals (de Leeuw, Schmid, & Mennen, 2010; de Leeuw, 2008). No differences between Dutch and English L2 environments were found. For L2 English speakers, ratings were positively correlated with age of emigration and, for L2 Dutch speakers, with the amount of contact with other L1 German speakers.

A subset of these data was used in another rating study (in which L2 speakers of German were also included). This time, attritors were not rated worse than controls (but L2 speakers were). Again, no differences between Dutch and English L2 environments were found. Ratings were not significantly correlated with length of residence or the amount of L1 use, but there was a significant positive correlation between native-likeness ratings and performance on an L1 C-test, a version of a fill-in-the-blanks test (Hopp & Schmid, 2013). Neither of these two studies investigated the possible segmental sources of the ratings of the bilingual speakers.

#### 1.3.2. Acoustical measurements

A comparatively large literature has focused on changes in segmental production in the L1 of bilingual speakers. The feature that has probably been studied most widely with respect to phonetic attrition is voice onset time (VOT) of voiceless stop consonants, such as /p, t, k/. VOT is the time between the burst of a stop consonant and the first vibration of the vocal folds in the following vowel. Languages are classified as short-lag or long-lag with the divide typically being set at 35 ms (Jansen, 2004; Keating, 1984; Lisker & Abramson, 1964). English and German are characterised as long lag, while many Romance languages (e.g., French, Spanish or Portuguese), but also some Germanic languages (e.g., Dutch) are short lag.

VOTs appear to be particularly susceptible to cross-linguistic assimilation. In an investigation of bilingual L1 speakers of English (L2 French) and L1 speakers of French (L2 English), the VOT of L1 voiceless stops of both groups was found to have changed towards the values of the other language in both their L1 and their L2, compared to a monolingual baseline. However, the attrition was only partial, as the L1 VOT did not reach the typical values of monolingual speakers of the respective L2 (Flege, 1987). Similar effects on L1 English have been found for L2 speakers of Brazilian Portuguese (Major, 1992) and Spanish (Williams, 1980). In a case study of two L1 Dutch monozygotic twin sisters, longer VOT values for voiceless stops were measured in the one sister who had been immersed in an L2 English environment for more than 30 years (Mayr, Price, & Mennen, 2012). But consonant articulation has been found to change even after much shorter periods of L2 immersion: In a longitudinal study of L1 English students, enrolled in a six-week beginners' Korean language course in Korea, the VOT of these speakers' English voiceless stops became gradually longer, approaching the longer VOT of Korean aspirated stops (Chang, 2012). Significant VOT changes were also attested in more experienced learners of Korean, but the effect was smaller than in beginning learners (Chang, 2013).

Other consonantal phenomena that have been investigated include the production of /l/ and /r/ in L1 German under the influence of L2 English: In an Anglophone Canadian immersion setting, L1 speakers of German were found to produce /l/ with a higher first formant (F1), characteristic of their L2. The second formant (F2), lower in North American English than in German, remained unchanged (de Leeuw, Mennen, & Scobbie, 2013; de Leeuw, 2008). In German natives with (rhotic) Irish English as their L2, formant changes indicative of rhoticity were found in word-final L1 /r/. Bilinguals exposed to (non-rhotic) British English, by contrast, showed no signs of rhoticity in their speech (Ulbrich & Ordin, 2014).

Articulatory changes in vowels through L2 exposure have received less attention in the literature than consonants: One study on L1 Russian/L2 English bilinguals, residing in the US, investigated the interface between consonants and vowels in the phonetic realisation of phonemic contrasts between syllable-final voiced and unvoiced consonants. It was found that bilinguals used the same strategies as monolinguals to realise the voicing contrast, but that they also made use of additional strategies that were not found in monolinguals. A positive correlation between L2 experience and deviation from the monolingual standard on vowel duration was established (Dmitrieva, Jongman, & Sereno, 2010).

Research on L2-induced vowel changes also includes the abovementioned studies on L1 English/L2 Korean learner groups. In these speakers, L1 vowel production was found to be influenced in terms of the fundamental frequency (f0) of vowels. Also, some drifts in the L1 vowel systems of especially female learners were evident (Chang, 2012). As in the case of VOT, changes were more pronounced in inexperienced than experienced learners (Chang, 2013). A study on two groups of L1 Korean speakers in an English immersion environment, by contrast, failed to find effects of L2 exposure on the L1 vowel system. This was true for speakers who had spent an average of 1.3 years in the US as well as for speakers who had been residing there for 8 years on average (Baker &

Trofimovich, 2005). A high daily share of L1 use in both groups (>40%) and low L2 proficiency in the short-residence group may have contributed to limiting the impact of the L2 immersion on L1 speech.

Most of these studies used elicited samples of isolated words or sentences for their analyses. This means that there is L2 influence even on careful, controlled L1 speech, so the effect has to be considered fairly robust. However, samples that have been elicited this way do not necessarily reflect spontaneous speech performance. It is furthermore difficult to judge in how far changes that have been measured would be noticeable by other native speakers. A case study of an L1 Dutch speaker who had been immersed in an Indonesian L2 environment for more than 30 years provides a window into the frequency of audible pronunciation changes. In one hour of spontaneous speech, only 48 instances of non-target-like pronunciations were found. The majority of these concerned stress and intonation patterns (Giesbers, 1997). No comparison to monolingual speakers was made and the ratings were impressionistic. The data therefore have to be interpreted with caution, but suggest that even after decades of L2 immersion, changes in the L1 are subtle.

#### 1.4. The present study

The link between global accent and specific articulatory changes in L1 attritors has received little attention so far. It is evident that immersion in an L2 setting can lead to a global accent on the one hand and to L2-related articulatory changes on the other. The main determinants of the changes of the global accent, however, still need to be identified. This study tries to take a step in this direction.

In the following, we report the results of three analyses of spontaneous speech data of functionally monolingual L1 speakers of German in Germany and bilingual L1 speakers of German in North America with English as their L2.

- 1) Native raters' perception of L1 speech was analysed to determine whether bilingual speakers are perceived as less native-like than speakers in a monolingual control group. On the basis of the native-likeness ratings of the bilinguals, speaker groups for the second study were formed.
- 2) A comparison of acoustical measurements on four speech sounds of German determined whether the production of these sounds can distinguish bilingual speakers who were perceived to sound like native speakers of their L1 from bilinguals who do not sound native-like. Both bilingual groups were also compared to a monolingual control group.
- 3) We attempted to establish a relationship between the rating results from the first study and the measurements from the second study within both bilingual groups.

### 2. Study 1: Native-likeness rating

#### 2.1. Hypotheses

Our hypotheses are based on the SLM, which states that diaphones gradually become more similar to one another in production. However, the model does not make specific predictions with respect to the global impact of such changes on the phonetic level. Still, it is conceivable that a multitude of drifts on the micro-level could lead to perceptible changes on the macro-level. Therefore, we hypothesise that bilingual speakers, as a group, differ from monolingual speakers and are perceived as less native-like. Earlier studies on attrition have singled out factors that are associated with attrition on the individual level: We followed Hopp and Schmid (2013) in their selection of three factors, namely length of residence in the L2 environment, self-reported L1 use and general L1 proficiency, as determined by a C-test. We expect these factors to be significantly associated with reduced L1 attrition in the sense that shorter length of residence, more L1 use and higher proficiency should be associated with more native-like ratings.

#### 2.2. Participants

In the rating study, we included recordings of two groups of speakers: (1) 33 L1 speakers of German, residing in Germany, who did not use any language other than German in their daily lives, and (2) 33 L1 speakers of German, residing in the USA or Canada, across a relatively small geographical distance in the New York and Toronto areas, where they made use of their L2, English, in their daily lives. We refer to the group of German residents as the rating control group and to the latter group of North-American residents as the rating attriter group.

The rating control group consisted of 22 female and 11 male speakers. Speakers originated from different regions of Germany, with a majority originating from or residing in the Frankfurt and Hamburg areas. The speakers were on average 38.18 years old (SD 11.14, range 22–65). Their average educational level was 4.06 (SD 0.93, range 2–5) on a scale from 1 to 5.<sup>1</sup> Speakers in both groups completed a German C-test, constructed by Schmid (2011), which was used to estimate their general proficiency in their L1, German. On this test, control group speakers scored an average of 92.86% correct responses (SD 3.27, range 81.4–97.7).

The rating attriter group consisted of 27 female and 5 male speakers. The speakers were on average 44.18 years old (SD 8.75, range 29–64). As determined by a Wilcoxon rank-sum test, they were significantly older than the control group ( $W=725$ ,  $p=0.02$ ).

<sup>1</sup> Education was coded on the following scale: 1=lower secondary education (German 'Volksschule/Hauptschule' or equivalents); 2=intermediate secondary education (German 'Realschule' or equivalents ≈ British O-level); 3=higher secondary education only (German '(Fach-)Abitur' or equivalents ≈ British A-level); 4=higher secondary education with additional vocational training; 5=higher secondary education with additional university degree.

With an average educational level of 4.45 (SD 0.67, range 3–5), attritors were more highly educated than controls, but this difference was not significant ( $W=666, p=0.09$ ). Attritors had emigrated to North America from all over Germany at an average age of 28.58 (SD 5.41, range 21–42). They had been residing in North America for 15.76 years on average (SD 7.79, range 6.5–34). Using a sociolinguistic questionnaire, we collected self-reported data of L1 use in different contexts (work, home and other): On average, attritors reported using their L1 for 20.72% of the time (SD 18.8, range 0–76.7). Participants in the rating attriter group also completed the proficiency test in German, scoring an average of 89.98% correct responses (SD 6.77, range 72.1–97.7), performing slightly less accurately than the rating control group speakers, but not significantly so ( $W=415, p=0.09$ ).

### 2.3. Materials

All speakers participated in a two-hour testing session that was part of a larger research project on bilingualism. In one of the tasks, they were asked to retell a ten-minute fragment of a Charlie Chaplin film that they had watched immediately before the retelling (Schmid, 2011). This task was preceded by a questionnaire interview (30–45 min) and an eye-tracking experiment (15–20 min; not reported here). During the entire testing session, participants were only addressed in German and only listened to German stimuli, which gave them 45–65 min to experience a monolingual L1 mode prior to the retelling task (Grosjean, 1982, 1998).

From each of the retelling recordings, an excerpt with an average length of 20 seconds was extracted (range 17–23 s). In selecting the excerpts, we took great care to ensure that they could not betray the linguistic background of the speaker in any way other than through the speaker's accent. The excerpts contained no English loan words, code switches or names. The excerpts also contained no errors and, as far as possible, no or few self-corrections, hesitations or long pauses. This was done to ensure that the raters would not mistake hesitancy, which might be caused by non-linguistic factors, for lack of spoken proficiency.

### 2.4. Method

The excerpts were rated by 30 native speakers of German (7 males), all students at the Johannes Gutenberg University of Mainz. Raters were on average 24 years old (SD 3.26, range 19–33). None of them reported having acquired any language other than German before entering elementary school and all reported that German was the language they spoke best.

The raters were presented with the excerpts in a web questionnaire that they filled in under the supervision of a German native speaker. The entire experiment was conducted in German. Every rater listened to all 66 recordings in the same pseudo-randomised distribution, in which samples from both speaker groups were evenly distributed, as were samples from male and female speakers. Two successive excerpts never described exactly the same movie fragment. After listening to each recording, raters answered two questions, given in English translation here: (1) Is this speaker a native speaker of German? Yes/No; (2) How certain are you of your judgement? Very certain/Somewhat certain/Uncertain. The answers were converted to a six-point Likert scale with the combination of the answers 'Yes' and 'Very certain' at one end of the scale (6) and the combination of the answers 'No' and 'Very certain' at the other (1). This method, pioneered by Moyer (1999), has been used in other studies on the speech of bilingual populations (e.g., de Leeuw et al., 2010; Schmid, Gilbers, & Nota, 2014). In a study on scale effects in accentedness ratings, this two-question approach was found to produce similar results to eliciting answers directly on a six-point Likert scale (Schmid & Hopp, 2014).

We analysed these data by means of a linear mixed-effects (LME) model (R, version 3.2.2; R Core Team, 2015). As a starting point, we constructed a model of the rating data that included all relevant fixed and random factors: Fixed factors were participant group, age, educational level, proficiency test score and, for the attritors, length of residence in the L2 environment and L1 use (Appendix 1, Model A). Of the latter five factors, the z-score was entered into the model. Random factors were speaker and rater. We constructed the optimal model by dropping fixed factors one by one and conducting formal model comparisons, excluding factors only when the comparison had shown that the more complete model was not superior to the reduced model. Complete model descriptions can be found in Appendix 1.

### 2.5. Results

On a six-point scale ranging from 1 (clearly non-native-like) to 6 (clearly native-like), speakers in the rating control group received an average rating of 5.05 (SD 0.61, range 3.7–6.0). Speakers in the attriter group received an average rating of 4.05 (SD 1.25, range 1.4–5.7). 13 of 33 attritors (39.4%) received a rating below the range that was defined by the rating control group speakers (i.e., below 3.7).

The optimal model (Appendix 1, Model B) included participant group and, for the attritors, L1 use and length of residence in the L2 environment as fixed factors. In this model, the intercept of 5.05 represents the rating that the average speaker in the control group received ( $SE=0.185, t=27.4, p<0.001$ ). In the attriter group, speakers received a rating that was lower by 1 point on average ( $SE=0.216, t=-4.66, p<0.001$ ). For the attritors, there was a significant effect of L1 use, with ratings increasing by 0.33 points for each standard deviation ( $SE=0.153, t=2.18, p=0.033$ ). Length of residence also had a significant effect, with ratings decreasing by 0.05 points for each additional year of residence ( $SE=0.02, t=-2.92, p=0.005$ ).

### 2.6. Discussion

Our hypothesis stated that bilingual speakers globally sound less native-like than monolingual speakers, presumably due to a multitude of small changes in the L1 phonological system that are induced by the L2. The data from the rating study confirm this

hypothesis, both on the individual and on the group level: Almost 40% of the attritors received lower ratings than the control group speaker with the lowest rating. As a group, attritors averaged one point below the control group on the six-point rating scale. Also, the attriter group exhibited larger variation than the monolingual group (4.3 vs. 2.3 points on the six-point scale).

From this, we conclude that immersion in an L2 environment had an impact on the attriter group as a whole, but that individuals were differentially affected. The effects of both length of residence in the L2 environment and L1 use went into the predicted direction: The longer the attritors lived in the L2 environment, the less native-like their L1 speech was rated, and the more they used their L1 in their daily lives, the more native-like their speech was perceived to be.

While our rating results do show that prolonged L2 exposure can have an impact on global accent, they cannot tell us what it is that bilinguals who sound like monolingual speakers do differently from bilinguals who do not sound like native speakers any more. In the following study, we investigate the differences between native-like and non-native-like bilinguals on the articulatory level. We focus on the production of four L1 speech sounds, comparing two groups of attritors to a monolingual control group.

### 3. Study 2: Acoustical measurements

#### 3.1. Hypotheses

In the detailed acoustical analysis of the speech samples, we expect that bilinguals who were perceived as native-like in the rating experiment (study 1) deviate less strongly from monolingual speakers in their sound production than non-native-like bilinguals.

If the predictions of the SLM are accurate, we should find that L1 German phonemes and their phonetic realisations shift in the direction of similar L2 English phonemes in attritors, who experience daily exposure to English.

To generate more specific hypotheses for the language combination discussed here, we based the selection of speech sounds to be analysed on the existing literature, that is, on a comparison of reference formant values of Standard German and North American English speech sounds. This is described below under [Section 3.3](#).

#### 3.2. Participants

In the measurements study, we included three groups of participants. These groups consisted of subsets of the two groups from the rating study—the control and the attriter group—and were selected as follows: First, all male participants were removed from the sample. Physical differences in the vocal tract between male and female speakers cause differences in vowel production (Rietveld & van Heuven, 2009). Mixed sex groups are therefore only recommended when the number of male and female speakers is balanced. Second, we removed all speakers who—as determined by the first author, a native speaker of German—did not speak Standard German, a supraregional variety widely used in formal and informal contexts in Germany (Krech, Stock, Hirschfeld, & Anders, 2009; Schmidt, 2005; Wiese, 1996). This concerned 5 of the 22 female speakers in the control group and 3 of the 27 female speakers in the attriter group. These speakers were excluded because no phonetic reference data are available for most non-standard varieties of German.

From the remaining 17 speakers in the control group, we selected 10 speakers, based on their results on the native-likeness rating (mean: 4.95), which were representative of the ratings control group as a whole (mean: 5.05). We refer to this group as the acoustical control group.

From the remaining 24 speakers in the attriter group, we selected two groups of 10 speakers each: The first group included speakers whose rating results (mean: 5.14) fell into the same range as the ratings of the 10 speakers in the acoustical control group. We refer to this group as the native-like attriter group, as it is representative of a subgroup of bilinguals who, after prolonged L2 immersion, still sound native-like to native raters. The second group included speakers whose rating results (mean: 2.74) all fell *below* the range defined by the ratings of the 10 speakers in the acoustical control group. We refer to this group as the non-native-like attriter group, as it is representative of a subgroup of bilinguals who, after prolonged L2 immersion, no longer sound native-like to native raters.

The differences between the groups with respect to the results of the native-likeness rating were significant, as determined by a Kruskal–Wallis test ( $\chi^2 = 19.8307$ ,  $df = 2$ ,  $p < 0.001$ ). The acoustical control group and the group of native-like attriters had equally high ratings, being statistically indistinguishable from one another ( $W = 37$ ,  $p = 0.34$ ). Non-native-like attriters had lower ratings than both other groups, differing significantly from both the acoustical control group ( $W = 100$ ,  $p < 0.001$ ) and the native-like attriter group ( $W = 100$ ,  $p < 0.001$ ).

Detailed participant characteristics of all three groups, including the rating results, and the results of group comparisons can be found in [Table 1](#). No additional significant differences between the groups were found in these comparisons.

#### 3.3. Materials

The comparison of the speech sound production was based on the proposed distinction between new, similar and identical phones (Flege, 1987, 1995). Mutual influence cannot occur in ‘new’ (existing in only one of the two languages of a bilingual) or ‘identical’ speech sounds, but only in ‘similar’ sounds. These are characterised by a comparable phonetic realisation in both languages on the one hand, making the speech sounds appear as cross-linguistic counterparts of one another, and articulatory

**Table 1**

Participant characteristics of the three groups in study 2.

	Acoustic control group (n=10)	Native-like attriter group (n=10)	Non-native-like attriter group (n=10)	Group comparison
Native-likeness rating	4.95 (SD: 0.43; 4.4–5.7)	5.14 (SD: 0.39; 4.4–5.7)	2.74 (SD: 0.64; 1.7–3.5)	$\chi^2=19.8307$ , df=2, $p<0.001$
Age	37.9 (SD: 8.03; 25–49)	45.8 (SD: 8.39; 33–62)	46 (SD: 9.99; 29–64)	$\chi^2=4.8506$ , df=2, $p=0.09$
Age of emigration	–	29.7 (SD: 4.88; 22–39)	25.6 (SD: 4.2; 21–33)	$W=69.5$ , $p=0.15$
Length of residence	–	16.6 (SD: 7.09; 10–31)	20.8 (SD: 9.11; 7–34)	$W=35$ , $p=0.27$
L1 use (self-reported)	–	31.5% (SD: 19.83; 6.7–76.7)	20.88% (SD: 20.93; 0–55)	$W=68.5$ , $p=0.17$
German proficiency test <sup>a</sup>	92.05% (SD: 4.38; 81.4–97.7)	89.77% (SD: 7.77; 72.1–97.7)	86.5% (SD: 6.16; 79.1–95.3)	$\chi^2=4.1241$ , df=2, $p=0.13$
Education <sup>b</sup>	4.0 (SD: 0.82; 3–5)	4.5 (SD: 0.71; 3–5)	4.3 (SD: 0.48; 4–5)	$\chi^2=2.5189$ , df=2, $p=0.28$

<sup>a</sup>Correct responses. Spelling errors were not counted as incorrect responses.<sup>b</sup>Self-reported on a scale from 1 to 5 (see footnote 1).

differences on the other, which, according to the SLM, could lead to association and convergence in the phonological system of a bilingual.

Our analysis focused on the values of the first (F1) and second formant (F2) of the speech sounds (Flege, Munro, & Fox, 1994). For selecting the German speech sounds that were to be analysed, we compared the average vowel formant values for female speakers of Standard German (Pätzold & Simpson, 1997) to the values of North American English (NAE). Of the three sources for NAE with separate data for female speakers, none was ideally suited for our purposes: The oldest data were collected from 48 speakers in Northern Central regions (Michigan and neighbouring states), which is reasonably close to the Toronto and New York areas where most speakers in the attriter group lived (Hillenbrand, Getty, Clark, & Wheeler, 1995). However, these measurements feature some peculiarities, including a reversal of the formant values of /æ/ and /ɛ/. Also, no long vowel in the low-back region—as an example of the lexical sets ‘palm’ or ‘start’ (Wells, 1982)—is included. A newer dataset, which does include the ‘start’ vowel, is based on recordings from 10 female speakers in the South and Southwest (Texas and neighbouring states), which is much farther away from the region where our data were collected (Yang, 1996). The ‘start’ vowel is also included in the most recent of the three sources: the *Atlas of North American English* (ANAE; Labov, Ash, & Boberg, 2005). We used a subset of the data, consisting of all five datasets from female speakers (2 from Toronto, 3 from New York City) that were available.<sup>2</sup> These monolingual data match our bilingual data in terms of regional background, but the number of speakers is lower than in the other two sources. In the following, we consider the data from all three sources.

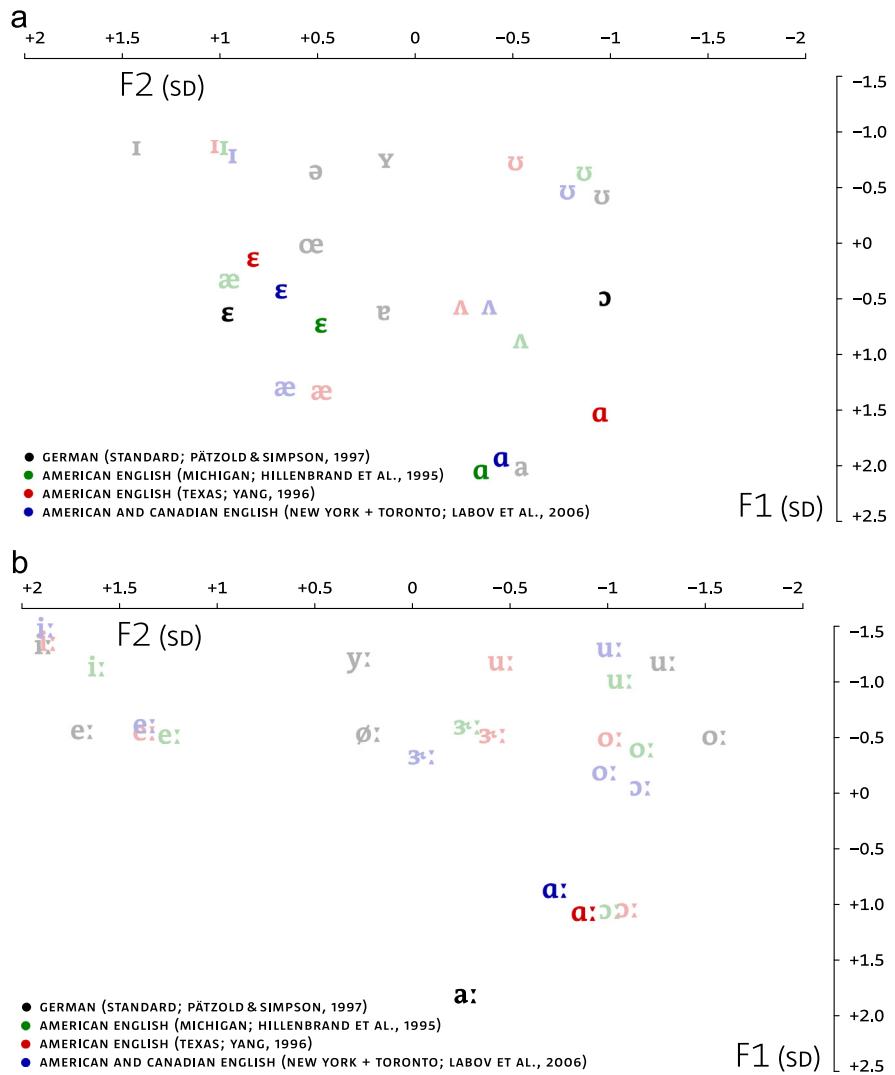
For cross-linguistic comparison, it is advisable to normalise measured values (in Hertz) of both vowel systems. We applied Lobanov's (1971) speaker-intrinsic normalisation method that has been shown to be one of the best in terms of preserving cross-linguistic differences (Adank, Smits, & van Hout, 2004; Flynn, 2011). The normalised sound systems of the two languages are shown in Fig. 1.

The selection of pairs of speech sounds in German and NAE that can be considered similar was based on three criteria, two of which are derived from factors known to affect segmental perception (Flege, 1988): (1) One traditional criterion for whether two speech sounds are similar is to determine if they are represented by the same IPA symbol in the phonetic literature. We used this as a first, rough guideline, eventually selecting only German speech sounds for which an English counterpart existed that was no further away than one standard deviation on the F1 or F2 scale in the normalised data. (2) The speech sounds also had to resemble one another in terms of lip rounding. (3) Apart from articulatory factors, we also considered orthographic effects: Our bilingual speakers are fully literate in both their languages and mostly started to acquire their L2 back in Germany in school. Through this type of instruction, which often relies heavily on written materials, they became aware of the sound-symbol correspondences in English and might have been influenced by them when integrating the L2 sounds into their phonological space. Also, we only included speech sounds that were sufficiently frequent in the recordings of the retellings. The vowels that were selected are /a:/ as in German *Bad* ‘bath’, /ɛ/ as in German *Bett* ‘bed’ and /ɔ/ as in German *Schrott* ‘junk’. Strange et al. (2007) found these to be clearly differentiated, non-overlapping categories of Standard German, which could mostly be reliably identified with their American English counterparts.

For German /a:/, the vowel /a:/ can be identified as the most similar speech sound in the NAE vowel system. While not always represented by the same IPA symbol, the two speech sounds are the only long, unrounded open vowels in the two phonological systems. Moreover, both are frequently represented by the same graphemes in their respective language (e.g., German *Vater* ‘father’ and English *father*). This speech sound is missing from the dataset in Hillenbrand et al. (1995). The two sets of reference data that contain this vowel (Labov et al., 2005; Yang, 1996) give similar formant values, indicating that the NAE sound is produced with a lower F1 and a lower F2, compared to the German sound. In the case of an ongoing merger of the two speech sounds, we expect the German sound to show lower F1 and F2 values in the attriter groups.

German /ɛ/ is phonetically most similar to NAE /ɛ/. Both speech sounds, represented by the same IPA symbol, are the only short, unrounded open-mid front vowels in the two systems. The speech sounds frequently occur in the same phonetic environments and are typically represented by the same graphemes (e.g., German *Bett* ‘bed’ and English *bed*). The three sets of reference data do not fully agree on the formant values of the corresponding English sound: All three give a lower F2 value, compared to the German sound. According to Yang (1996) and Labov et al. (2005), the F1 value should also be (slightly) lower than in German. Based on

<sup>2</sup> We are grateful to Daniel Ezra Johnson for sharing the formant data with us.



**Fig. 1.** F1/F2 formant values of Standard German and North American English short vowels (top panel; 1a) and long vowel (bottom panel; 1b) from reference datasets, normalised using the speaker-intrinsic method described in [Lobanov \(1971\)](#). Vowels in a lighter shade were not included in the present study.

Hillenbrand et al. (1995), by contrast, no change (or even a slight rise) in the F1 would be predicted. The authors have pointed out that these values for /æ/ and /ɛ/ are at odds with older measurements (Hillenbrand, 2003) and have to be treated with some caution.

For the third speech sound we selected, German /ɔ/, it was most difficult to identify an English equivalent. The low-back part of the North American vowel system does not feature a vowel that is described as short and rounded. One candidate for influence on the German sound is the vowel in words like 'pot', which our reference sources transcribe as /ɑ/, that is, as an unrounded vowel. This speech sound is short, like German /ɔ/, and we see an overlap in the graphemic representation of the sounds (e.g., German *Pott* 'pot' and English *pot*). Our reference data agree on the F1 value of /ɑ/, which is higher than that of the German speech sound, but disagree on the F2: [Yang \(1996\)](#) gives a value that is similar to that of the German sound, but the other two sources indicate higher F2 values. A second candidate for influencing the production of the German sound would be /ɔ:/, which was identified as the American English counterpart of German /ɔ/ by [Strange et al. \(2007\)](#). The two speech sounds match in terms of roundedness, but less so in terms of length and graphemic representation (e.g., English *caught*). The reference sources again do not fully agree on the formant values: All three suggest that the F2 value of /ɔ:/ is similar, or slightly lower, than that of the German sound. The F1 of the English speech sound, by contrast, is lower in the data from the ANAE, but lower by about the same value in the two other datasets. Summing up, we expect the German sound to shift towards a higher F1, which characterises both /ɑ/ and /ɔ:/, according to the majority of the sources (with the exception of the data for /ɔ:/ from the ANAE). For the F2, either a slight lowering or a substantial rise can be expected, depending on the source and on the sound that bilinguals identify as an equivalent of German /ɔ/.

In addition to the three vowels, we analysed one consonantal speech sound, the lateral approximant /l/ in syllable onset position. Its production has been studied before in L1 German/L2 English bilinguals, revealing that /l/ qualifies as an indicator of phonetic attrition (de Leeuw, 2008). The production of /l/ differs between Standard German and NAE insofar as the English sound is produced with a higher F1 and lower F2 (Ladefoged & Maddieson, 1996; Recasens, 2004; Wells, 1982).

### 3.4. Method

For all 30 speakers, the first and second formant (F1 and F2) of 20 instances of each of the four phonemes were measured in PRAAT (Boersma & Weenink, 2012). The sound samples were drawn from the entire recording in all cases and all speech sounds were measured in stressed syllables. As the data originate from spontaneous speech, the control over speech production is limited, compared to elicitation studies. The exact lexical production contexts for all 20 instances of each speech sound could not be matched across all speakers. Within the given constraints, all possible care was taken to ensure the reliability of the current measurements. Where possible, vowels were measured in plosive contexts, as these contexts have the least amount of coarticulation effects on the vowel. No vowels followed by a liquid were measured, as liquids are known to produce large coarticulation effects. Formant measurements were made at the intensity peak (or dip, for onset /l/), as these generally occur near the midpoint of the sound.

Our formant data were normalised and analysed by means of a linear mixed-effects (LME) model (R, version 3.2.2). Fixed factors were participant group, speech sound and formant as well as z-scores for age, educational level, the score on the German proficiency test and, for the attritors, length of residence in the L2 environment and self-reported L1 use (Appendix 2, Model A). As before, we constructed the optimal model by successively excluding factors that did not improve the model. Complete model descriptions can be found in Appendix 2.

### 3.5. Results

The average formant values of all speech sounds analysed here are given in Table 2 and visualised in Fig. 2.

The optimal model (Appendix 2, Model B) included age as a fixed factor ( $\beta = -0.018$ ,  $SE = 0.007$ ,  $t = -2.55$ ,  $p = 0.016$ ), in addition to interactions of participant group, speech sound and formant. Based on this, pairwise comparisons between the participant groups were calculated for each sound and formant using Tukey's HSD. The results of these comparisons for each speech sound are reported below. To avoid Type 1 errors, we applied FDR (False Detection Rate) correction to the p-values.

#### 3.5.1. /a:/

Based on the reference values for /a:/, the phonetically most similar American English speech sound, we expected L2 influence on the German sound to lead to lower F1 and F2 values.

For the F1, native-like attritors had a slightly higher group mean than the control group, whereas non-native-like attritors had a slightly lower mean than the controls. None of the group differences were significant, neither between controls and native-like attritors ( $z = 1.52$ ,  $p = 0.39$ ) or non-native-like attritors ( $z = 0.06$ ,  $p = 0.94$ ) nor between the two attriter groups ( $z = -1.54$ ,  $p = 0.39$ ).

For the F2, both groups of attritors had lower values than controls, with native-like attritors deviating more strongly from the controls than non-native-like attritors. The difference between the controls and the native-like attritors was significant ( $z = -3.13$ ,  $p = 0.02$ ), whereas non-native-like attritors did not differ significantly from controls ( $z = -1.75$ ,  $p = 0.39$ ) or native-like attritors ( $z = 1.45$ ,  $p = 0.4$ ).

#### 3.5.2. /ɛ/

The phonetically most similar English speech sound, /ɛ/, should—in the case of L2 influence on L1 production—lead to lower F2 values in the German sound. For F1, no definite hypothesis could be formulated, given that two sources cited slightly lower values (Labov et al., 2005; Yang, 1996) and one gave a slightly higher value (Hillenbrand et al., 1995).

On average, the F1 values in both groups of attritors hardly differed from those in the control group. The pairwise comparisons showed that none of the groups differed from any other (controls vs. native-like attritors:  $z = 0.51$ ,  $p = 0.86$ ; controls vs. non-native-like attritors:  $z = 0.16$ ,  $p = 0.95$ ; native-like vs. non-native-like attritors:  $z = -0.37$ ,  $p = 0.95$ ).

The mean F2 values for the two attriter groups were almost identical and both lower than those of the controls. However, there were no significant differences between controls and native-like attritors ( $z = -1.53$ ,  $p = 0.39$ ) or non-native-like attritors ( $z = -1.6$ ,  $p = 0.39$ ). The difference between the two attriter groups was not significant either ( $z = -0.08$ ,  $p = 0.95$ ).

#### 3.5.3. /ɔ/

Two speech sounds of American English, /a/ and /ɔ:/, have been identified as potential sources of influence on the German sound. According to the majority of our reference sources, interference from both speech sounds would lead to a higher F1 value in German.

**Table 2**

Acoustical data of the three groups in study 2 for the four German sounds.

Phoneme	Formant	Acoustic control group	Native-like attriter group	Non-native-like attriter group
/a:/	F1	776.3 (SD: 144.33)	791.25 (SD: 99.3)	760.59 (SD: 99.3)
	F2	1552.31 (SD: 228.41)	1416.60 (SD: 142.05)	1463.32 (SD: 180.77)
/ɛ/	F1	595.78 (SD: 105.59)	590.75 (SD: 100.70)	584.31 (SD: 85.78)
	F2	2094.43 (SD: 221.59)	1998.40 (SD: 225.01)	1994.45 (SD: 239.27)
/ɔ/	F1	603.91 (SD: 115.12)	610.64 (SD: 86.42)	593.02 (SD: 87.56)
	F2	1179.02 (SD: 197.39)	1127.27 (SD: 154.06)	1144.75 (SD: 158.03)
/l/	F1	397.95 (SD: 97.55)	459.34 (SD: 135.80)	442.44 (SD: 104.60)
	F2	2125.11 (SD: 296.25)	2066.1 (SD: 302.95)	2001.38 (SD: 340.07)

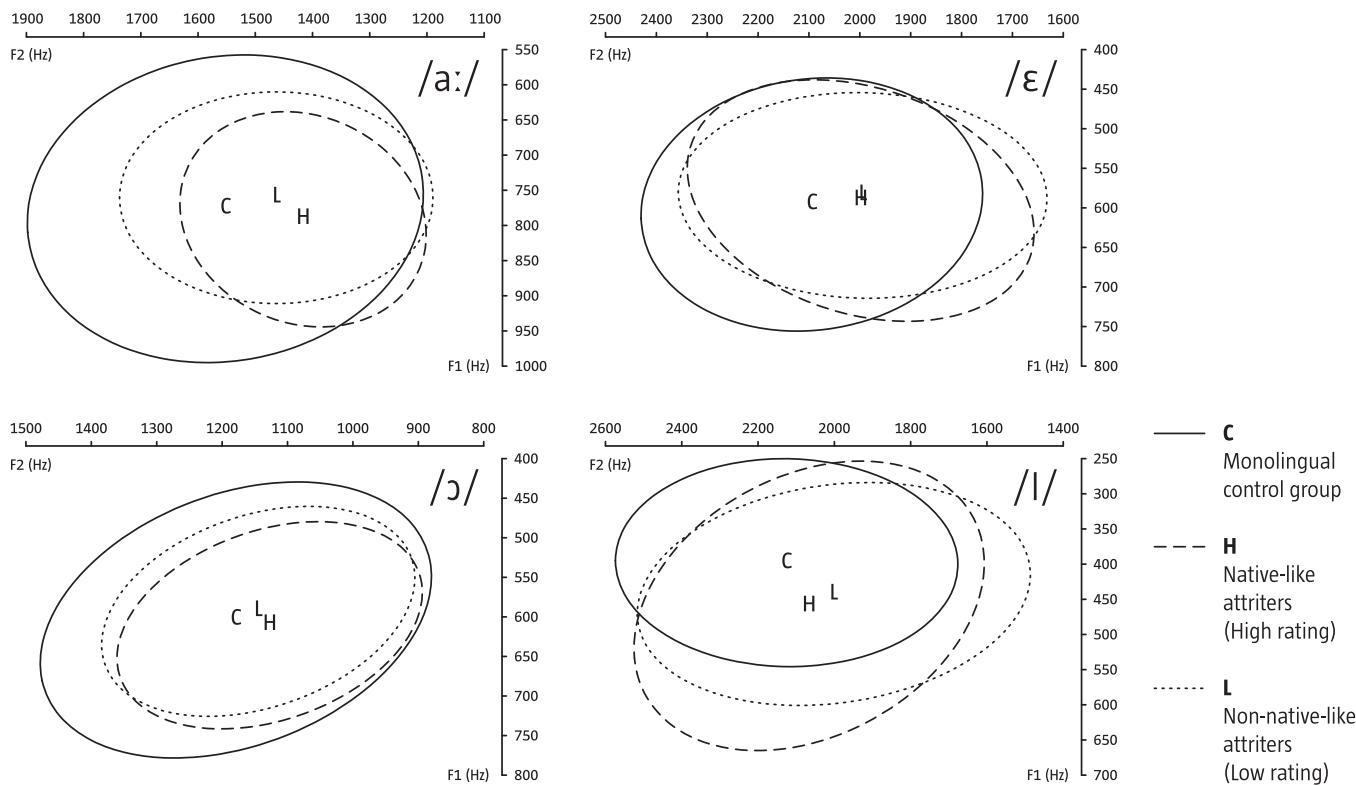


Fig. 2. F1/F2 formant values of all four sounds in study 2. Letters indicate the mean for each of the three groups. Circles indicate one standard deviation around the group mean.

The reference sources did not agree on the F2 of the English speech sounds, predicting either a slight rise or a substantial lowering of the L1 value.

On the F1, the groups only differed minimally with slightly higher values in the native-like attriters and slightly lower values in the non-native-like attriters than in the controls. There were no significant differences between controls and native-like attriters ( $z=1.19$ ,  $p=0.47$ ) or non-native-like attriters ( $z=0.19$ ,  $p=0.95$ ). The attriter groups did not differ significantly from one another either ( $z=-1.06$ ,  $p=0.54$ ).

For the F2, both attriter groups had slightly lower mean values than controls. Neither native-like attriters ( $z=-0.9$ ,  $p=0.63$ ) nor non-native-like attriters ( $z=-0.3$ ,  $p=0.95$ ) differed significantly from control group speakers. Additionally, there was no reliable difference between the two attriter groups ( $z=0.63$ ,  $p=0.8$ ).

### 3.5.4. //

The possible influence of American English // on the equivalent German speech sound could be reflected in a higher F1 and a lower F2 in the production of the bilingual speakers.

The mean F1 values in both attriter groups were higher than in the control group with native-like attriters having the highest value of all three groups. The difference from the control group was significant for both attriter groups (native-like:  $z=3.78$ ,  $p=0.003$ ; non-native-like:  $z=2.97$ ,  $p=0.024$ ). There was, however, no significant difference between the attriter groups ( $z=-0.84$ ,  $p=0.64$ ).

For the F2, the values in both attriter groups were on average lower than in the control group with non-native-like attriters having the lowest value. Neither of the two attriter groups differed significantly from the control group (native-like:  $z=-0.25$ ,  $p=0.95$ ; non-native like:  $z=-1.38$ ,  $p=0.4$ ). The difference between the attriter groups was not reliable either ( $z=-1.19$ ,  $p=0.47$ ).

## 3.6. Discussion

For this study, we formulated a twofold hypothesis: First, we predicted that the results of the rating study, in which some attriters were judged as sounding non-native-like, would be mirrored in the acoustical analysis on the segmental level. On the group level, we expected speakers who sounded native-like to show less or no deviance from a monolingual control group in their production of three vowels and one consonant, whereas attriters who sounded non-native-like were expected to deviate more from monolingual controls. Second, we predicted on the basis of the SLM that sound changes in attriters, if detectable, should be directed towards similar speech sounds in the phonological system of the L2, English.

The first hypothesis was not confirmed by the data: There was no formant on which the non-native-like attriters differed from the controls while native-like attriters did not differ from the controls. Rather, we saw two cases that point in the opposite direction: First, native-like attriters had a significantly lower F2 value than controls for /a:/, whereas non-native-like attriters did not. In the latter group, a tendency towards lower F2 values was also evident, but was less pronounced than in the native-like attriters and non-significant.

Second, native-like attritors had higher F1 values than non-native like attritors for /ɪ/; that is, the native-like attritors were farther removed from the controls than the non-native-like attritors. However, both attriter groups differed significantly from controls on this formant.

From this, we conclude that differences in the segmental production of the four speech sounds under investigation were not the main factor that made attritors sound non-native-like. It is still interesting to see that in some of the speech sounds, at least one of the attriter groups deviated from the monolingual norm.

The second hypothesis—a shift of the L1 formant values towards those of similar L2 speech sounds—received more support from our data: Native-like attritors had lower F2 values for /a:/, compared to the controls, which made this vowel more similar to its English diaphone. This was also true for the F1 of /ɪ/, which is higher in North American English than in German and was also found to be higher in both attriter groups, compared to the controls. We conclude that the L2, while apparently having a similarly strong impact on both attriter groups, seems to have acted as one of the factors of influence on some speech sounds of the L1 system of the bilingual speakers. However, the number of significant differences between the monolingual group and the bilingual groups was fairly low. In two of the speech sounds, the vowels /ɛ/ and /ɔ/, neither of the formants differed between groups. Of these, /ɔ/ is the more interesting case: We saw that the lower back part of the English vowel system does not feature a vowel that matches the quantity and quality of the German sound (see [Section 3.3](#)). As potential sources of influence, /a/ (as in 'pot') and /ɔ:/ (as in 'caught') were identified. It is, however, conceivable that neither of the L2 speech sounds is sufficiently similar to the L1 sound to be seen as its diaphone by bilingual speakers. In this case, no merged L1/L2 category would be created and no L2 influence on the L1 would occur. It also has to be noted that /ɛ/ and /ɔ/ are the two speech sounds for which our reference sources for North American English gave divergent information. Such regional differences in the production of these vowels may be perceived by bilingual speakers as well. Even if our participants permanently live in one region, they are likely to be exposed to regional varieties of English from all over North America. The absence of a clear English point of reference may also stand in the way of identifying L2 sounds as diaphones of L1 sounds.

Participant characteristics—such as length of residence in the L2 environment, L1 use, education or general L1 proficiency—did not have significant effects on the formant values of the speech sounds we have measured. This may partly be due to the fact that there were no significant differences between controls and attritors on the majority of the formants, that is, there was no non-native-like speech production in the bilingual group that would need to be explained by these factors.

#### 4. Study 3: Relationship between rating results and acoustical measurements

##### 4.1. Hypotheses

If the articulation of L1 segments is one of the main sources for a global non-native-like accent, it should be possible to establish a relationship between the ratings (study 1) and the articulatory deviance from monolingual speakers on individual speech sounds (study 2). We emphasise that isolated significant effects cannot be interpreted in favour of a general relationship of the two variables. Only a widespread pattern of effects in the same direction would constitute evidence of a direct link between global ratings and individual articulation.

##### 4.2. Participants

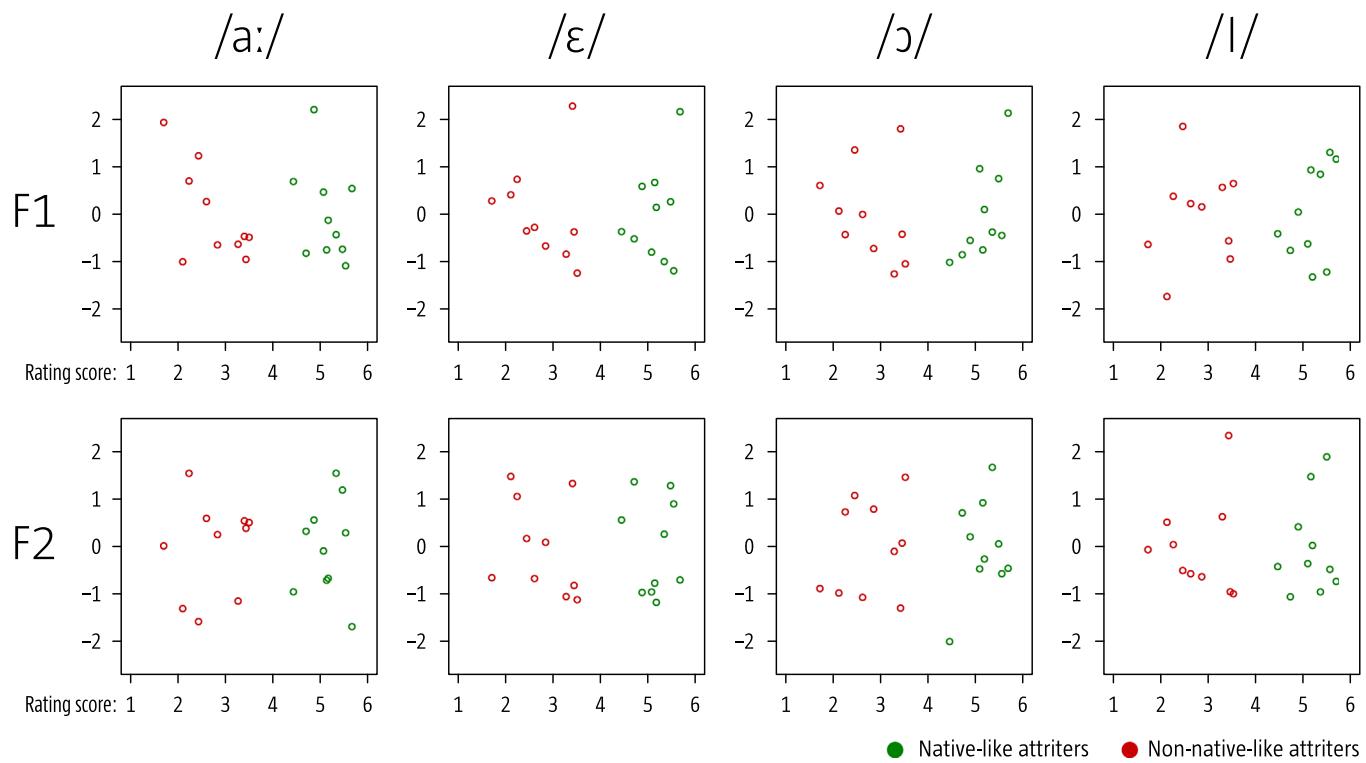
The participants in this study were identical to the speakers in the two attriter groups—native-like and non-native-like—from study 2.

##### 4.3. Materials

We combined the results of the second study, the acoustical measurements, with the results of the first study, the native-likeness ratings. For each speech sound and formant, we calculated the difference between the mean value of all acoustical control group speakers and the mean value for each individual attriter, thereby quantifying the difference between each bilingual speaker and an average monolingual speaker. We then converted the difference values into their absolute values (i.e., non-negative values). This was done because there is no reason to assume that native-likeness ratings would be influenced by the *direction* of the deviation from control group speakers, rather than by its *strength*. Consistently producing a speech sound with a formant value 200 Hz above the average native value is just as likely to sound non-native-like as producing the same sound with a formant value 200 Hz *below* the monolingual standard.

##### 4.4. Method

As before, we used a linear mixed-effects (LME) model (R, version 3.2.2) to analyse the results of the native-likeness ratings. We followed the same procedure as before: Starting from a model that included z-scores for age, education, proficiency test score, L1 use and length of residence in the L2 environment as fixed factors as well as speaker and rater as random factors ([Appendix 3](#), Model A), we tried to construct the optimal model for the rating data by excluding factors that did not improve the model. The data in this analysis are not identical to the data in study 1, as the present analysis only includes a subset of the attriters, namely of the speakers whose recordings were used for acoustical measurements in study 2.



**Fig. 3.** Speaker averages of the rating results and of the normalised absolute difference values for each sound and formant. For the formant values, the difference between the speaker average of each bilingual attriter and the group average of the monolingual control group was calculated in order to quantify the acoustical deviation in the bilingual speakers (cf. Section 4.3 for more detail).

The optimal model for only the rating data (Appendix 3, Model B) included, in addition to participant group, length of residence in the L2 environment as a fixed factor. In this model, the intercept was 5.06 (SE=0.185,  $t=27.41$ ,  $p<0.001$ ). Significant effects were found for participant group ( $\beta=-2.24$ , SE=0.197,  $t=-11.37$ ,  $p<0.001$ ), reflecting the lower rating of the non-native-like group, and length of residence ( $\beta=-0.035$ , SE=0.012,  $t=-2.88$ ,  $p=0.01$ ), reflecting a lower rating for participants who had been living in the L2 environment for longer. We compared this optimal model to models that additionally included the normalised absolute difference values of the first and second formants of all four speech sounds (Appendix 3, Model C). Complete model descriptions can be found in Appendix 3. We have visualised the average rating values and the average normalised absolute difference values for each speaker in Fig. 3.

#### 4.5. Results

The optimal model for the rating data (Appendix 3, Model D) included, in addition to participant group and length of residence, the normalised absolute difference values of the F1 of /a:/, /e/, /o/ and /i/ and of the F2 of /e/ and /o/. Compared to a model without the acoustical data, this model was found to be superior ( $\chi^2=37.52$ ,  $p<0.001$ ). For the native-like attriter group, significant effects of acoustical data on the rating results were evident for only two out of eight formants: The difference values for the F1 of /o/ ( $\beta=0.38$ , SE=0.121,  $t=3.15$ ,  $p=0.002$ ) and /i/ ( $\beta=0.28$ , SE=0.126,  $t=2.22$ ,  $p=0.027$ ) were positively related with the rating results, that is, speakers whose vowel production differed more strongly from that of an average control group speaker tended to receive higher ratings.

For the non-native-like attriter group, there were significant acoustical effects on the rating data for four out of eight formants: A negative relationship between the rating results and the vowel production was evident for the F1 of /a:/ ( $\beta=-0.76$ , SE=0.117,  $t=-6.48$ ,  $p<0.001$ ) and /e/ ( $\beta=0.88$ , SE=0.258,  $t=3.39$ ,  $p=0.001$ ) as well as for the F2 of /e/ ( $\beta=-0.88$ , SE=0.212,  $t=-4.13$ ,  $p<0.001$ ), that is, speakers who differed less strongly from the control group tended to receive higher ratings. For the F2 of /o/, there was a positive relationship between rating results and acoustical data ( $\beta=0.53$ , SE=0.223,  $t=2.37$ ,  $p<0.018$ ).

#### 4.6. Discussion

Based on the hypothesis that articulatory changes in individual speech sounds translate to global accent changes, we investigated the relationship between the results of the rating study and the acoustical measurements. We expected ratings to be negatively associated with formant differences (i.e., smaller distances to the control group should be associated with higher ratings in individual speakers).

In the native-like attriter group, a majority of the formant data showed no significant effects of the acoustical data on the rating results. The effects that we saw did not point in the predicted direction. Rather, there was a positive association between the variables, which is difficult to interpret: the larger the deviation from the control group, the higher the perceived native-likeness.

In the non-native-like attriter group, the effects of four formants were significant in our model. In three out of these four cases, there was a negative association between the variables, as our hypothesis predicted: the smaller the deviation from the control group, the higher the perceived native-likeness. In the fourth case, the association was positive (i.e., larger deviation was associated with higher ratings). This means that even in the non-native-like group, most formants did not produce any effect in the direction we predicted. Taken together, there was no evidence for a widespread pattern of negative relationships between rating results and acoustical data, as predicted by our hypothesis.

A possible explanation for these findings is suggested by Strange, Levy, and Law (2009), who investigated the perceptual assimilation of L1 German vowels by naïve monolingual American English listeners in elicited speech. For the /ɛ/ vowel, for instance, their results showed that most listeners perceived German /ɛ/ as a good exemplar of American English /ɛ/, irrespective of large F1 differences between the languages. This can be explained, they claim, by Selective Perceptual Routines (SPRs; Strange, 2006), tying in with current models of speech perception (Best, 1995; Flege, 1995, 2002; Kuhl, 1993, 2000). These “highly over-learned” L1 SPRs “enable the listener to extract the most reliable phonologically-relevant information rapidly from the incoming speech stream in order to recover the intended message” (Strange et al., 2009, p. 1462). In this model, listeners assimilate speech sounds to expected categories, with low levels of attention for details that are unnecessary to derive meaning from the utterance. As this effect was found for the perception of short stretches of elicited speech, it stands to reason that it could be even stronger in the present study, in which raters listened to 20 s of spontaneous speech.

## 5. General discussion

This study has focused on the phonetic changes in L1 speech that prolonged exposure to an L2 can cause. Flege's *Speech Learning Model* (SLM) assumes that the phonetic categories of the L1 and L2 of a bilingual speaker exist in one common phonological space. In this space, mutual influence between the speech sounds of both languages is predicted. More specifically, the SLM hypothesises that similar sounds of the L1 and L2 are linked to one another, on the basis of individual perception, in one representation. In speech sounds that are associated across languages, so-called diaphones, the SLM predicts category assimilation, that is, the sounds come to resemble one another over time.

We investigated these predictions using spontaneous speech samples of monolingual and bilingual L1 speakers of German. The bilingual speakers (L1 attritors) had emigrated from Germany to North America, where they spoke English as their L2. In a first step, we assessed the global changes in the accent of both speaker groups by means of a rating study in which other L1 speakers of German judged the native-likeness of speech samples. It was found that the L1 attritors, as predicted, differed from the monolingual controls: As a group, attritors were rated one point lower on a six-point scale than controls. Also, their ratings exhibited a wider range than those of controls (4.3 vs. 2.3 points). The observation that bilingual groups exhibit more interpersonal variation even in their L1 has been made before and has been attributed to the increased diversity in linguistic input that bilinguals experience, leading to a larger number of potential factors influencing the speech behaviour of these individuals (de Leeuw et al., 2013).

Almost 40% of the speakers in the attriter group received an average rating below the monolingual range. From this, it is clear that immersion in an L2 environment did not make all speakers in our attriter group sound non-native-like in their L1, but that a sizeable minority of the speakers differed from the monolingual controls in how native-like they were perceived to be. Given that different levels of L1 and L2 exposure constitute the only systematic difference between controls and attritors, it seems reasonable to attribute the rating difference to the immersion setting in which the attritors have been living. The negative relationship between the rating results and length of residence in the L2 environment and the positive association between the ratings and L1 use point in the same direction.

In a second step, we attempted to trace the physical origin of the global differences in the ratings by conducting an analysis of the first and second formants of four speech sounds—three vowels and one consonant—of German. Three groups of ten speakers each were included in the analysis: monolingual controls, attritors who were rated identical to the control group and attritors who were rated as sounding non-native-like. Based on the SLM, it was predicted that articulatory differences would only be found between non-native-like attritors and control group speakers, whereas native-like attritors, who are perceptually indistinguishable from controls, should pronounce the sounds the same as controls. The SLM also hypothesizes that speech sounds in the L1 of bilingual speakers would come to resemble equivalent L2 sounds, the two speech sounds being linked as diaphones.

The expectation that native-like and non-native-like attritors differ in their production of L1 speech sounds was not confirmed by the data: Both bilingual groups differed from the monolingual control group on at least one formant. However, native-like attritors differed from controls on more formants and, where both groups differed from controls, more strongly than non-native-like attritors. Given these results, we conclude that the differences in perceived non-native accent between the two attriter groups cannot be accounted for by how they pronounced the four speech sounds under scrutiny. It is still interesting to note that L2 exposure, as the only systematic difference between controls and attritors, had an impact on the pronunciation of a small number of speech sounds in bilinguals in both groups.

The second hypothesis of the SLM, namely that the formants of L1 speech sounds would come to resemble those of similar L2 sounds, was supported: Whenever significant differences between attritors of either group and controls were found, the German speech sounds shifted in the direction of their most similar English counterparts. These findings contribute to a small body of evidence for L2-induced changes in L1 vowel systems. Our results are also in line with a previous investigation of the articulation of German /i/ by attritors in Anglophone immersion: These speakers, like ours, produced a significantly higher F1, as is typical for their

L2. For F2, we found a lower value in attritors relative to controls, a tendency that was also evident in previous studies (de Leeuw et al., 2013; de Leeuw, 2008).

While the changes that we observed are systematic, the underlying predictors remain unclear: Participant characteristics, such as L1 use, length of residence or general L1 proficiency, were not found to contribute significantly to the models of our acoustical data from any of the bilingual groups. Also, there was no robust evidence for a negative relationship between articulatory differences and ratings.

In summary, the two hypotheses of the SLM received some support from our data: First, we were able to show that the sustained presence of an L2 in the phonological system of a speaker can lead to changes in the articulation of L1 speech sounds. It has to be kept in mind, however, that 60% of the attritors performed no worse than functionally monolingual natives. This echoes the conclusions drawn from earlier investigations, namely that phonetic attrition does not necessarily affect each speaker and that, even in the speakers whose speech has come to deviate, changes are subtle rather than dramatic (Giesbers, 1997; Hopp & Schmid, 2013). Our data—in particular the relationship between the rating results and the length of residence in the L2 environment as well as L1 use—cannot fully explain why some speakers deviated from the monolingual norm. Data on variables such as language aptitude and task switching ability could help elucidate factors involved in the preservation of a native-like pronunciation in the L1. Language aptitude in particular is one of the factors that have been assumed to mediate the success of both L1 maintenance and L2 acquisition in late bilinguals (Abrahamsson & Hyltenstam, 2008; Bylund, Abrahamsson, & Hyltenstam, 2012), but this relationship has not yet been demonstrated in the domain of phonetics.

Second, our results show that, as the SLM predicts, L2-oriented changes in L1 pronunciation can occur. These changes, though, did not differentiate the attritors who were perceived as native-like in their L1 from attritors who were not, at least not as far as the speech sounds investigated in this study are concerned. The present data therefore cannot tell us the source of the perceived non-native accent in our bilingual speakers. Our analyses centred on qualitative effects, but differences in vowel quantity may also be brought about by attrition. This aspect, as well as the interaction of vowel quantity and quality, has received little attention so far. The same holds for sound shifts involving diphthongs. An advantage of spontaneous speech samples is that they are closer to natural speech production than more systematically elicited data, but on the downside, only a limited number of speech sounds occur in all recordings with sufficient frequency for acoustical analyses. This was one of the reasons that our study focused on monophthongic phonemes.

Future directions include investigations of whether the specific acoustical differences that were found in some speech sounds are sufficiently salient to be perceived by monolingual listeners. Such analyses, combining spontaneous and elicited speech, and including general cognitive data, will offer further insight into the sources and nature of non-native-like accents in the L1.

## Acknowledgements

This research was supported by a grant (016.104.602) of the Netherlands Organisation for Scientific Research (NWO), awarded to Monika S. Schmid.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.wocn.2016.07.001>.

## References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), 481–509, <http://dx.doi.org/10.1017/S027226310808073X>.

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

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, 48(1), 1–27.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, 161, 25–35, <http://dx.doi.org/10.1016/j.actpsy.2015.07.015>.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Boersma, P., & Weenink, D. (2012). PRAAT: Doing phonetics by computer. Retrieved from [www.praat.org](http://www.praat.org).

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1–L2 encoding of manner in speech and gesture. *Studies in Second Language Acquisition*, 30(02), 225–251, <http://dx.doi.org/10.1017/S0272263108080327>.

Brown, A., & Gullberg, M. (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English. *Bilingualism: Language and Cognition*, 14(01), 79–94, <http://dx.doi.org/10.1017/S1366728910000064>.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language?. *Studies in Second Language Acquisition*, 34(2), 215–241, <http://dx.doi.org/10.1017/S0272263112000034>.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268, <http://dx.doi.org/10.1016/j.wocn.2011.10.007>.

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), 520–533, <http://dx.doi.org/10.1016/j.wocn.2013.09.006>.

de Leeuw, E. (2008). *When your native language sounds foreign: a phonetic investigation into first language attrition* (PhD thesis). Edinburgh: Queen Margaret University.

de Leeuw, E., Mennen, I., & Scobbie, J. M. (2013). Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism*, 17(6), 683–700, <http://dx.doi.org/10.1177/1367006912454620>.

de Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, 13(1), 33–40, <http://dx.doi.org/10.1017/S136672890990289>.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll, & A. M.B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 179–201). Oxford: Oxford University Press.

Dmitrieva, O., Jongman, A., & Sereno, J. (2010). Phonological neutralization by native and non-native speakers: the case of Russian final devoicing. *Journal of Phonetics*, 38(3), 483–492, <http://dx.doi.org/10.1016/j.wocn.2010.06.001>.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: a review* (pp. 224–401). Norwood, NJ: Ablex.

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. (2002). Interactions between native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: papers in honor of Henning Wode* (pp. 217–244). Trier: Wissenschaftlicher Verlag.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34(2), 153–175, <http://dx.doi.org/10.1016/j.wocn.2005.05.001>.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95(6), 3623–3641, <http://dx.doi.org/10.1121/1.409931>.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40(4), 467–491.

Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series*, 2(11), 1–28.

Giesbers, H. W. M. (1997). Dutch in Indonesia: language attrition or language contact?. *Dutch Overseas: Studies in Maintenance and Loss of Dutch as an Immigrant Language*, 163–188.

Grosjean, F. (1982). *Life with two languages: an introduction to bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, F. (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131–149.

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics*, 17(1), 53–78, [http://dx.doi.org/10.1016/S0911-6044\(03\)00054-X](http://dx.doi.org/10.1016/S0911-6044(03)00054-X).

Gürel, A. (2008). Research on first language attrition of morphosyntax in adult bilinguals. *Second Language Research*, 24(3), 431–449, <http://dx.doi.org/10.1177/0267658308093611>.

Hillenbrand, J. (2003). American English: Southern Michigan. *Journal of the International Phonetic Association*, 33(1), 121–126, <http://dx.doi.org/10.1017/S0025100303001221>.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111. Retrieved from (<http://www.ncbi.nlm.nih.gov/pubmed/7759650>).

Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: the impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 34(2), 361–394, <http://dx.doi.org/10.1017/S0142716411000737>.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: a laboratory phonology approach to English, Hungarian, and Dutch* (PhD thesis). Groningen: University of Groningen.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60(2), 286–319, <http://dx.doi.org/10.2307/413642>.

Keijzer, M. (2010). The regression hypothesis as a framework for first language attrition. *Bilingualism: Language and Cognition*, 13(1), 9–18, <http://dx.doi.org/10.1017/S1366728909990356>.

Köpke, B. (2002). Activation thresholds and non-pathological first language attrition. In F. Fabbro (Ed.), *Advances in the neurolinguistics of bilingualism: essays in honor of Michel Paradis* (pp. 119–142). Udine: Forum.

Krech, E.-M., Stock, E., Hirschfeld, U., & Anders, L. C. (2009). *Deutsches Aussprachewörterbuch*. Berlin: de Gruyter.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychologica*, 128(3), 416–430, <http://dx.doi.org/10.1016/j.actpsy.2008.02.001>.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2), 119–135, <http://dx.doi.org/10.1017/S1366728906002483>.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: the Native Language Magnet theory. In B. Boysson-Bardies, S. Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: speech and face processing in the first year of life* (pp. 259–274). Dordrecht: Springer, <http://dx.doi.org/10.1007/978-94-015-8234-6>.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11850–11857, <http://dx.doi.org/10.1073/pnas.97.22.11850>.

Labov, W., Ash, S., & Boberg, C. (2005). *The Atlas of North American English*. Berlin: de Gruyter.

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

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49(2), 606–608.

Mägiste, E. (1979). The competing language systems of the multilingual: a developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior*, 89, 79–89.

Major, R. C. (1992). Losing English as a first language. *Modern Language Journal*, 76(2), 190–208.

Major, R. C. (2010). First language attrition in foreign accent perception. *International Journal of Bilingualism*, 14(2), 163–183, <http://dx.doi.org/10.1177/1367006910363063>.

Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch–English bilinguals: the case of monozygotic twin sisters. *Bilingualism: Language and Cognition*, 15 (4), 687–700, <http://dx.doi.org/10.1017/S136672891100071X>.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, 21, 81–108, <http://dx.doi.org/10.1017/S0272263199001035>.

Newport, E. L., Bavelier, D., & Neville, H. J. (2011). Critical thinking about critical periods: perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain, and cognitive development: essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

Opitz, C. (2010). L1 attrition and L2 acquisition: global language proficiency and language dominance in adult bilinguals. *EUROSLA Yearbook*, 10(1), 248–281, <http://dx.doi.org/10.1075/eurosla.10.13opi>.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: theoretical perspectives* (pp. 121–133). Amsterdam: John Benjamins.

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. In A. P. Simpson, K. J. Kohler, & T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech: Acoustic base, processing tools and analysis results* (pp. 215–247).

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics*, 11(2), 175–205.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from ([www.r-project.org](http://www.r-project.org)).

Recasens, D. (2004). Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control. *Clinical Linguistics & Phonetics*, 18(6–8), 593–603, <http://dx.doi.org/10.1080/02699200410001703556>.

Ribbert, A., & Kuiken, F. (2010). L2-induced changes in the L1 of Germans living in the Netherlands. *Bilingualism: Language and Cognition*, 13(1), 41–48, <http://dx.doi.org/10.1017/S1366728909990320>.

Rietveld, A. C. M., & van Heuven, V. J. (2009). *Algemene fonetiek*. Bussum: Coutinho.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436, <http://dx.doi.org/10.1006/jpho.1997.0051>.

Schmid, M. S. (2002). *First language attrition, use and maintenance: the case of German Jews in anglophone countries*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/sibil.24>.

Schmid, M. S. (2004). First language attrition: the methodology revised. *International Journal of Bilingualism*, 8(3), 239–255, <http://dx.doi.org/10.1177/13670069040080030501>.

Schmid, M. S. (2011). *Language attrition*. Cambridge: University Press.

Schmid, M. S. (2014). The debate on maturational constraints in bilingual development: a perspective from first-language attrition. *Language Acquisition*, 21(4), 386–410, <http://dx.doi.org/10.1080/10489223.2014.892947>.

Schmid, M. S., Gilbers, S., & Nota, A. (2014). Ultimate attainment in late second language acquisition: phonetic and grammatical challenges in advanced Dutch–English bilingualism. *Second Language Research*, 30(2), 129–157, <http://dx.doi.org/10.1177/0267658313505314>.

Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: range and rater effects. *Language Testing*, 31(3), 367–388, <http://dx.doi.org/10.1177/0265532214526175>.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 17(4), 729–748, <http://dx.doi.org/10.1017/S1366728913000771>.

Schmidt, J. E. (2005). Die deutsche Standardsprache: eine Varietät – drei Oralisierungsnormen. In L. M. Eichinger, & W. Kallmeyer (Eds.), *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?* (pp. 278–305). Berlin: de Gruyter.

Schmitt, E. (2010). When boundaries are crossed: evaluating language attrition data from two perspectives. *Bilingualism: Language and Cognition*, 13(1), 63–72, <http://dx.doi.org/10.1017/S1366728909990381>.

Strange, W. (2006). Second-language speech perception: the modification of automatic selective perceptual routines. *Journal of the Acoustical Society of America*, 120(5), 3137, <http://dx.doi.org/10.1121/1.4787743>.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naïve American listeners. *Journal of the Acoustical Society of America*, 126(3), 1461–1476, <http://dx.doi.org/10.1121/1.3179666>.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *Journal of the Acoustical Society of America*, 122(2), 1111–1129, <http://dx.doi.org/10.1121/1.2749716>.

Tsimpli, I. M., Sorace, A., Heycock, C., & Filippi, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257–277, <http://dx.doi.org/10.1177/13670069040080030601>.

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, 45, 26–42, <http://dx.doi.org/10.1016/j.wocn.2014.02.008>.

Weinreich, U. (1957). On the description of phonetic interference. *Word*, 13, 1–11.

Wells, J. C. (1982). *Accents of English, Vol. 3: beyond the British Isles*. Cambridge: University Press.

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

Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Vol. 2: perception*, Vol. 2 (pp. 185–215). New York: Academic Press.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24(2), 245–261, <http://dx.doi.org/10.1006/jpho.1996.0013>.