The language dependency of /m/ in native Dutch and non-native English

Boer, M. de; Heeren, W.F.L.

OCTOBER 09 2023

The language dependency of /m/ in native Dutch and non-native English

Meike M. de Boer ; Willemijn F. L. Heeren

J. Acoust. Soc. Am. 154, 2168–2176 (2023)
https://doi.org/10.1121/10.0021288

JASA ARTICLE

The language dependency of /m/ in native Dutch and non-native English

Meike M. de Boer[a] and Willemijn F. L. Heeren
Leiden University Centre for Linguistics, Leiden University, Reuvensplaats 3-4, 2311 BE Leiden, The Netherlands

ABSTRACT:
In forensic speaker comparisons, the current practice is to try to avoid comparisons between speech fragments in different languages. However, globalization requires an exploration of individual speech features that may show phonetic consistency across a speaker's languages. We predicted that the bilabial nasal /m/ may be minimally affected by the language spoken due to the involvement of the rigid nasal cavity in combination with a lack of fixed oral articulatory targets. The results show that indeed, L1 Dutch speakers ($N = 53$) had similar nasal formants and formant bandwidths when speaking in their L2 English as in their native language, suggesting language-independency of /m/ within speakers. In fact, acoustics seemed to rely more on the phonetic context than on the language spoken. Nevertheless, caution should still be exercised when sampling across languages when the languages' phoneme inventories and phonotactics show substantial differences. © *2023 Acoustical Society of America*.
https://doi.org/10.1121/10.0021288
(Received 7 February 2023; revised 11 September 2023; accepted 14 September 2023; published online 9 October 2023)
[Editor: Ewa Jacewicz] Pages: 2168–2176

I. INTRODUCTION

More than four out of five citizens in the Netherlands speak at least one foreign language (Eurostat, 2021). Consequently, speech evidence in criminal cases is collected in multiple languages, sometimes even within one recording (Van der Vloed *et al.*, 2014; cf. Campbell *et al.*, 2004). Thus, whereas the majority of academic research in forensic speech science is performed in a monolingual context (cf. Mok *et al.*, 2015, but see Lo, 2021), the practice requires investigating speaker-specificity from a multilingual perspective. More specifically, there is a need to explore language-independent speech characteristics within speakers (cf. Bradlow *et al.*, 2017). In principle, a language mismatch is expected to deteriorate speaker classification (cf. Ma and Meng, 2004), because within-speaker variation between the samples increases. Hence, such a mismatch in casework is advised against the code of practice of the International Association for Forensic Phonetics and Acoustics (IAFPA, 2020). However, some individual speech features may show phonetic consistency across a speaker's languages when they occur similarly in both languages involved in a forensic speaker comparison so that they could be selected to be used in auditory-acoustic forensic speech analysis. One such candidate is the bilabial nasal /m/.

In a prior study, we investigated whether filled pauses may be language-independent for speakers of English as a second language (L2) and Dutch as a first language (L1; de Boer and Heeren, 2020). The filled pauses *uh* and *um* are produced relatively unconsciously (as argued by e.g., Clark and Fox Tree, 2002), are phonetically quite similar in Dutch and English, and are often not taught explicitly in language classes. This led to several linguists predicting that they may be transferred from the L1 and thus be language-independent (e.g., Clark and Fox Tree, 2002; de Leeuw, 2007). We showed that speakers tended to adapt their first and second vowel formants (F1, F2) to the target language rather than transferring them from their native language. Whereas other phonetic parameters (e.g., F3, fundamental frequency) appeared to remain consistent across the two languages, the study showed that even a feature that is thought to be produced largely unconsciously can be language-dependent within speakers.

The current study investigates the language-dependency of the speech sound /m/ in L1 Dutch and L2 English among a group of Dutch speakers who learned English as an L2. Between-language differences may occur when a speaker mentally stores an L2 segment in a different phonetic category than an L1 segment (see, e.g., de Bot, 1992; Flege *et al.*, 2021). As the bilabial nasal /m/ is among the most similar phonemes when comparing native Dutch and English (Collins and Mees, 2003), with no apparent *a priori* differences in production, between-language differences within a speaker at this level are unexpected. In Sec. I A, we discuss phonological/phonotactic and acoustic properties of /m/, and in I.B we discuss /m/ acoustics from a cross-linguistic perspective.

Due to the involvement of the rigid nasal cavity, /m/ acoustics largely depend on stable anatomical properties—the nose is not an active articulator—and effects of language-dependent articulatory settings may be limited as well. In addition, because of this, /m/ is argued to be speaker-specific in one's native language (e.g., Glenn and Kleiner, 1968; Fujimura, 1962; Rose, 2002). Since its

[a]Electronic mail: m.m.de.boer@hum.leidenuniv.nl

https://doi.org/10.1121/10.0021288

acoustics may be largely language-independent, whereas it contains speaker-specific information, it may be a useful speech feature in cross-language forensic settings.[1] In addition, by using a within-speaker design to compare acoustics across two languages, the effect of language can be studied in the absence of physical and anatomical differences between speakers, providing a more transparent picture of cross-linguistic differences (cf. Ng *et al.*, 2012). As argued below, if the acoustics of /m/ turn out to be language-dependent in L1 Dutch and L2 English, this would be most likely a property of the general articulatory settings, which would imply that language-independent segments may be hard or even impossible to find. If, however, we find no between-language differences, this could mean that the language spoken may be less impactful in the segment's acoustics than speaker-specific properties.

In addition to multilingual forensic speaker comparisons, a more thorough insight into language-specificity can be useful for linguistic areas dealing with L2 teaching and speech production. Regarding L2 teaching, knowledge of how speakers do or do not adjust their speech production helps in identifying bottlenecks in learning a particular language, which could be addressed in pronunciation training. Subtle differences are often neglected in L2 teaching, especially in languages with relatively similar phonetic structures such as English and Dutch. In addition, regarding (language-specific) speech production, not much is known regarding how different articulatory settings may affect the production and acoustics of individual speech sounds. Although Dutch and English are relatively similar languages, which may lead to less within-speaker variability across languages, they show pronunciation differences and differences in their phoneme inventories (e.g., English /θ, ð, ae/ and Dutch /ɤ, x, ɛi/, see Collins and Mees, 2003). In addition, nasals are quite comparable in other languages as well, making our study potentially relevant for different language combinations.

A. The bilabial nasal /m/

In both Dutch and English, the bilabial nasal /m/ is a common phoneme, claimed to be produced similarly in both languages (Collins and Mees, 2003; Tobias, 1959). In addition, in both languages, the nasal inventory consists of the same three phonemes—the bilabial /m/, the alveolar /n/, and the velar /ŋ/—which occur in similar phonetic contexts and positions (Collins and Mees, 2003). As the alveolar nasal, the bilabial nasal can occur in prevocalic, intervocalic, and postvocalic positions. Hence, speakers of both Dutch and English are expected to store the phoneme /m/ in one shared category for both languages (e.g., Flege *et al.*, 2021).

Nasal consonants are produced with a lowered velum, creating resonance in the nasal cavity (e.g., Stevens, 1998, pp. 187–193). Due to a constriction somewhere along the oral tract, air is blocked and solely released through the pharyngeal-nasal passage. The nasal cavity lengthens the vocal tract by on average 12.5 cm (for adult males; Johnson, 2003, p. 152) and adds soft tissue, standing air, and nostril hairs, which absorb energy. This damping broadens the bandwidth of nasal consonants when compared to oral segments (Fant, 1970), and leads to a lower amplitude and resonance frequencies (Smorenburg and Heeren, 2021). Although the nasal and pharyngeal cavities form the main acoustic pathway (Tabain *et al.*, 2016), the oral cavity is involved and functions as a side-branch, absorbing even more energy and producing antiresonances (e.g., Stevens, 1998; Tabain *et al.*, 2016). In the production of /m/, the oral cavity is closed off by the lips (Fujimura, 1962), which leads to a relatively long side-branch (for men: 8 – 9 cm; Stevens, 1998), whereas the tongue is assumed to have no specific position (Su *et al.*, 1974).

Although much work has been done on distinguishing place of articulation for the different nasals, empirical descriptions of spectral characteristics are limited (cf. Tabain *et al.*, 2016; Smorenburg and Heeren, 2021). Overall, acoustic models agree that nasals have a low first formant of 200 – 300 Hz (e.g., Fujimura, 1962; Johnson, 2003; Stevens, 1998; Tabain *et al.*, 2016), which may overlap with the fundamental frequency (*F*0) due to the wide bandwidth (around 100 Hz; Stevens, 1998, pp. 187 – 194). A second nasal formant (N2) is estimated at around 1000 Hz, a third (N3) around 2200 Hz, and a fourth (N4) around 2800 – 3000 Hz. The N1 and N3 are associated with resonances in the pharynx cavity, and the N2 and N4 with resonances in the nasal cavity (e.g., Tabain *et al.*, 2016), of which the N4 specifically with the nasal system above the uvula (Fant, 1970).

Acoustic descriptions specific to /m/ are especially scarce, and a much-cited reference is based on one male speaker of Russian. This male's nasal formants (N1–N4) were around 150 Hz and 1, 2, and 3 kHz (Fant, 1970), but of course, measurements may vary somewhat for different speakers. Tabain *et al.* (2016, p. 896) measured /m/ acoustics for 21 (mostly female) speakers of three Australian languages and found formants (N1–N4) around 300 – 350, 1400, 2600, and 3700 Hz. Smorenburg and Heeren (2021) found somewhat lower formants for Dutch males in telephone speech, measured over the 800–3400 Hz band: mean values for nasal formants were found at 1067 (N2), 2035 (N3), and 2717 (N4) Hz.[2]

B. Variation within and between speakers

Because of its voiced and nasal nature, /m/ is among the most speaker-specific consonants, with relatively high between-speaker variability and low within-speaker variability (e.g., Glenn and Kleiner, 1968; Fujimura, 1962; Kavanagh, 2012; Rose, 2002; Schindler and Draxler, 2013; Van den Heuvel, 1996). Within-speaker variability of /m/ is described as lower than that of phonemes produced orally because the nasal cavity is relatively rigid compared to the oral cavity (Rose, 2002) and the nose does not actively move during the period when there is oral closure (Glenn and Kleiner, 1968, p. 368). Unless the speaker has a cold or

has (had) an injury or surgery affecting the nasal cavity, acoustics relying on the nasal cavity show relatively low variability within an individual over time. Thus, /m/ is often used in casework (Gold and French, 2011), although nasal acoustics are weak, which makes them difficult to measure in the low quality (telephone) recordings that are often used in forensic settings (Smorenburg and Heeren, 2021).

A source of within-speaker variability is related to a sound's phonetic context: the acoustics of /m/ can be affected by adjacent speech sounds (e.g., Fujimura, 1962; Kurowski and Blumstein, 1987). During the production of /m/, the tongue has no distinctive target and is free to move towards its place for the following vowel, which may lead to anticipatory coarticulation (Su *et al.*, 1974, p. 1877). Fant (1970, p. 147) describes that nasals may be especially coarticulated with adjacent back vowels, leading to a contracted pharynx and a shift in the higher nasal formants. Indeed, Smorenburg and Heeren (2021) found that preceding and following back-articulated contexts affected /m/ acoustics. When the following phonetic neighbor (consonant or vowel) had a back place of articulation, N2, N4, and the bandwidth of N4 (BW4) decreased and BW3 increased. When /m/ was preceded by a back-articulated neighbour, N3 and BW3 increased and BW2 and BW4 decreased. Overall, Smorenburg and Heeren (2021) concluded that back contexts have a lowering effect on nasals' resonance frequencies and that for /m/, this is especially visible in back contexts following the nasal onset position.

Between-speaker variability of /m/ is relatively high because of anatomical differences between speakers in the internal structure of the nasal cavity in addition to that of the oral cavity (e.g., Rose, 2002). According to Stevens (1998, pp. 19 – 20), there is a large variation between speakers in the volumes of their nasal cavities; even the left and right passages of one individual are often asymmetrical in size and shape. In addition, speakers differ in the timing of the opening and closure of the velopharyngeal port (Stevens, 1998, pp. 487 – 488).

The question remains whether the acoustics of the bilabial nasal are language-dependent. As stated above, descriptions of language-specific acoustics of /m/ are sparse in the literature, and the acoustic descriptions of a Russian individual from Fant (1970) are often cited as general (i.e., language-independent) spectral characteristics of /m/. According to Stevens (1998, pp. 487–488), languages may differ in the timing of the closure and release of the bilabial nasal, which is supported by a study on velopharyngeal timing in nasalised vowels (Solé, 1995). Tabain *et al.* (2016, p. 898) studied nasals in mostly female speakers of three Australian languages. Although they do not report on cross-linguistic differences, their figures show that there may be some differences between the languages: approximately 50 – 250 Hz for the nasal formants (N1–N4) and up to 110 Hz for the four bandwidths (BW1–BW4). Note, however, that the sample sizes per language were very low ($N = 5 – 9$). The study by Smorenburg and Heeren (2021) on /m/ acoustics among Dutch males reported formants that were much lower than those in Tabain *et al.* (2016). Despite a difference in biological sex for the participants used in both empirical studies, their findings indicate that for at least some parameters, languages may differ in /m/ acoustics.

A point of caution when comparing /m/ in different languages using a between-subject design is that its acoustics are related to the speakers' anatomy, as discussed in Sec. I A. Differences found between languages may thus be attributed to anatomical differences between the speakers on individual or group levels, e.g., different nasal structures between Aboriginal peoples and Western Europeans (cf. Ohki *et al.*, 1991), rather than be an attribute of the languages itself. It remains unclear whether (closely related) languages show cross-linguistic differences in /m/ acoustics, and whether differences found between the studies discussed above are in fact language-dependent rather than speaker-dependent. Thus, language-dependency can best be studied in a within-subject design, where the nasal cavity and other anatomical and physical features remain constant. To our knowledge, this has not yet been described in the literature.

Even when the Dutch and English /m/ are indeed stored in a shared phonemic category, different acoustic properties of /m/ may still be introduced when a speaker adapts their articulatory setting as a whole to the target language or if the phonotactics, i.e., nature and/or distribution of neighboring sounds (e.g., Weber and Cutler, 2006), are different between the L1 and L2. According to Collins and Mees (2003, Chap. 21), in RP English, the lips are looser and the tongue and throat muscles are more relaxed than in Standard Dutch. In addition, they describe that the tongue tip tends to be raised and active in RP English, whereas it is usually inactive in Dutch. Due to differences in phonotactics or sound co-occurrences, languages differ in coarticulation patterns, where the amount of coarticulation has been linked to the extent to which distinctions have to be made with other sounds (see, e.g., Lubker and Gay, 1982; Solé, 1992). As described previously, nasal inventories in Dutch and English are similar, and /m/ can be found in similar phonetic contexts. Hence, overall, limited language-dependency is expected—although the prevalence of preceding and following back phonetic neighbors may be different, and differences in overall articulatory settings may affect the acoustics of individual segments.

C. Research question and predictions

The current study investigated the within-speaker consistency of /m/ in a cross-linguistic context, comparing nasal acoustics in L1 Dutch and L2 English. The research question is whether /m/ acoustics remain consistent within the same speakers across the two languages, or whether they are language-dependent. If the bilabial nasal is language-independent within speakers, this would show its relative robustness against speaker-external factors, hence supporting its reputation as a potentially good speaker discriminator (e.g., Kavanagh, 2012). In addition, it would be a first step

https://doi.org/10.1121/10.0021288

towards attempting forensic speaker comparisons with materials in more than one language.

Because Dutch and English /m/ have similar phonetic-acoustic properties and occur in similar phonetic positions, we expect that Dutch L2 speakers of English have one single mental category for /m/ in both languages, which should lead to language-independency of /m/ acoustics. Due to the high involvement of the rigid nasal cavity, we predict that general acoustic settings have limited effect on its production and /m/ acoustics are largely language-independent, i.e., remain consistent within the same speakers. As the N2 and N4 are particularly associated with the nasal cavity, these measurements are expected to show even less variation across languages than N1 and N3, which are related to the pharynx cavity. Regarding bandwidths of the nasal formants, which are highly linked to the amount of soft tissue hence damping on the nasal, we expect speakers to remain consistent across languages as well.

However, although within-speaker consistency is overall expected to hold across languages, we know that the rigid nasal cavity does not prevent nasal acoustics from being affected by coarticulation (e.g., Su *et al*., 1974). This means that there may be differences in /m/ acoustics due to subtle differences in phonotactics, i.e., the frequencies of particular sound combinations in the two languages. As /m/ tokens with back-pronounced acoustic neighbours have different acoustics, e.g., lowered N2 (Smorenburg and Heeren, 2021), and /m/ tokens may be found more often in back contexts in one language than the other, such phonotactic differences may lead us to find a language effect due to coarticulation effects rather than actual cross-linguistic differences. To account for this possible confound, we include information on the tokens' left (i.e., preceding) and right (i.e., following) phonetic neighbor in the analysis, i.e., whether they are back or non-back (cf. Smorenburg and Heeren, 2021). We expect that although /m/ acoustics may be context-dependent, this does not alter their predicted within-speaker consistency across languages. Hence, when using /m/ in forensic speaker comparisons, sampling tokens from similar contexts may be more important than sampling them from speech in the same language.

Although we do not expect specific language-dependent acoustic differences for the bilabial nasal, what we do know is that articulatory settings as a whole may differ by language. The differences in articulatory settings between Dutch and English have no predictable effect on the nasal cavity itself but may affect the oral cavity thus influencing nasal acoustics.

II. METHODS

A. Materials and procedure

For this study, we used spontaneous speech recordings from the Database of the Longitudinal Utrecht Collection of English Accents (D-LUCEA; Orr and Quené, 2017). This database consists of recordings from the international student population of University College Utrecht (UCU), a liberal arts and sciences college where English is the lingua franca. For students who spoke English as a second language (L2), parts of the recording were also done in their native language (L1). This makes it a convenient database to study the same speakers in more than one language. As L2 English proficiency is one of the requirements to be selected to the highly competitive UCU, the level of English of the students can be described as at least above-average (Quené *et al*., 2017) and is estimated to be above B2 level according to the Common European Framework of Reference for Languages (see Council of Europe, 2019). Being situated in the Netherlands, about 60% of UCU's student population has Dutch as their L1, who are also overrepresented in the corpus and were selected as the focus of the current study.

The recording sessions consisted of several speech tasks, of which most of them in (L2) English. After a couple of reading tasks, each student was asked to give a two-minute informal monologue in their L1, in this case, Dutch, and then immediately in L2 English,[3] to end with a formal monologue and interactive part in English with an interlocutor who could understand both Dutch and English. For the current study, we used informal monologues, since this was the only task that was available in two languages for all students and has higher external validity than read speech. During recruitment, participants were informed about the 2-min monologues, which resulted in different levels of preparation including no preparation at all and merely selecting a topic beforehand. Topics of the informal monologues included hobbies, vacations, and the introduction period of UCU. Overall, the monologues can be described as (semi-)spontaneous (cf. Orr and Quené, 2017).

The sessions were recorded by eight microphones, of which we selected the close-talking microphone that was attached to a headset (Sennheiser HSP 2ew, Sennheiser, Wedemark, Germany). Further details about the recording conditions are described by Quené *et al*. (2017).

B. Speaker characteristics

For this study, the largest possible homogenous group of speakers was selected from the database, with the requirement that they were recorded within one month after their arrival at UCU. Recordings that were made at a later stage were potentially affected by convergence in the L2 (see Orr and Quené, 2014) and assimilation in the L1 (see Chang, 2012). This resulted in a group of 53 female speakers who spoke Dutch as their sole native language and without a salient regional accent. The students were $17-20$ years old (mean, $M=18.4$; standard deviation, $SD=0.8$) and had learned English as an L2 at a Dutch high school. In the Netherlands, English is taught as a second language from the final stages of primary school (around 11 years old). In addition, some students described having been on an exchange to an English-speaking country. Although all students were able to express themselves fluently in English, also demonstrated by the lack of an increase in hesitation

markers in the L2 (see de Boer and Heeren, 2020), some had a more audible Dutch accent than others. Although in the Netherlands, most schools use British English as a model (Chen, 2009), classes are often taught by Dutch L2 speakers of English and learners have a lot of American English input through pop culture (Smakman and De France, 2014). Hence, there is no clear target variety, which was also reflected in the students' self-reports (cf. de Boer and Heeren, 2020).

C. Segmentation and measurements

/m/ tokens were identified in Praat (Boersma, 2001) based on hand-made orthographic annotations with the aid of the Praat automatic alignment feature. All segment boundaries were corrected manually by repeated listening in combination with visual inspection of the spectrogram, focusing on its low N1 and relatively high distance between the N1 and other formants (Fujimura, 1962). This led to 3081 /m/ tokens in total (14–56 tokens per speaker per language). Per token, the neighbouring vowels or consonants were annotated and categorised as back or non-back. In addition to back vowels (/u, ɔ, o, a, ɑ/), the back category consisted of consonants with velar to uvular place of articulation (/k, g, ŋ, x, χ/); the non-back category consisted of all other contexts including silent pauses (cf. Smorenburg and Heeren, 2021). Table I shows the number of tokens for each of the contexts per language.

We used Praat (Boersma, 2001) to measure the first four nasal formants (N1–N4) and their bandwidths (BW1–BW4) in the middle 50% of the /m/ tokens. The measurements were done over the $0-5000$ Hz band using the Burg method, querying four formants. There were no undetected values, but values >2 SD from the grand mean were treated as outliers ($n=94$ for N1; 154 for N2; 156 for N3; $n=89$ for N4).

D. Statistical analysis

Based on our predictions and previous analyses with *lme4* (Bates *et al.*, 2015), we expected null results (see de Boer and Heeren, 2021). Hence, we built Bayesian linear mixed-effects models (LMMs), which allow evaluation of the strength of evidence in favor of or against the null hypothesis (see, e.g., Nicenboim and Vasishth, 2016). LMMs were built in R (R Core Team, 2022) using the *brms* package (Bürkner, 2017, 2018). Building on the script by Hugo Quené as used in de Boer *et al.* (2022), we used NUTS sampling (Bürkner, 2018) to sample prior distributions 44 000 times (in four independent chains of 11 000 samples each, of which the first 1000 warmup samples were discarded, so that 40 000 samples remained). The prior distributions for N1−N4 were based on acoustic-phonetic theory on the bilabial nasal, which is modelled by Stevens (1998) as approximately $250-300$, 1250, 2150, and 3050 Hz for males. For our female speakers, priors for N1−N4 were estimated at 350, 1300, 2200, and 3100 Hz (3.5, 10.2, 13.7, 15.9 Bark), respectively. For the bandwidths BW1−BW4, we used 100, 330, 340, and 350 Hz (0.8, 3.3, 3.4, 3.5 Bark) to estimate the priors (cf. Tabain *et al.*, 2016). For each parameter, 1000 prior predictive distributions were sampled and plotted in a scattergram with their means and SDs. These results confirmed the validity of the chosen priors.

Bayesian LMMs were built including random intercepts and random slopes for speakers. The models included the contrast-coded (−0.5, 0.5) fixed effects Language (Dutch, English), Left and Right Context (back, non-back), and the interactions of the Contexts with Language: Left*Lang and Right*Lang.[4] The fixed parts were first investigated using the 95% highest density interval (HDI); this is the smallest interval containing 95% of the posterior distribution. When this interval contains zero (0), this means that no significant difference was found between the levels of a factor. In addition, we calculated Bayes Factors (BF) using the *bayestestR* package (Makowski *et al.*, 2019). For the interpretation of the BFs, we used the Andraszewicz *et al.* (2015) adjusted version of Jeffreys (1961), as presented in Table II. In general, a BF of 1 indicates that there is no evidence in favor of or against H0, $BF>1$ indicates evidence against the H0, and $BF<1$ presents evidence in favor of H0.

III. RESULTS

Of the eight acoustic measurements included in this study, the HDIs showed that there was no main effect of Language for any of the models except for BW1 (see Table III). The absence of a Language effect was explored further using BFs (see Table IV). The BFs indicate that there is no difference between Dutch and English /m/ acoustics measured in this study, including BW1. We found evidence in favor of H0 for all acoustic parameters, which was very strong for N1 and N3; strong for N2, N4, BW3, and BW4; moderate for BW2; and anecdotal for BW1. The lack of a language effect is reflected in the overall means in L1 Dutch

TABLE I. Number of /m/ tokens per context and language (Dutch: 1759; English: 1322). The total number of tokens is 3081.

	Dutch (L1)		*English (L2)*	
Context	*left*	*right*	*left*	*right*
non-back	1,152	1,401	944	1,052
back	607	358	378	270

TABLE II. Overview of the interpretation of BFs (adopted from Andraszewicz *et al.*, 2015, p. 526). $BF>1$ gives evidence against H0; $BF<1$ gives evidence in favour of H0.

$BF>1$	$BF<1$	strength of evidence
>100	<0.01	extreme
30 – 100	0.01–0.033	very strong
10 – 30	0.033–0.1	strong
3 – 10	0.1–0.33	moderate
1 – 3	0.33–1	anecdotal

https://doi.org/10.1121/10.0021288

TABLE III. Means (and HDIs) for (a) the N1−N4 models and (b) the BW1−BW4 models (in Bark). Factor levels: Language (Dutch, English); Left and Right Context (back; non-back, NB). Factors with HDIs including 0 did not contribute to the optimal model and are indicated in gray.

(a)		*N1*	*N2*	*N3*	*N4*
Intercept	Mean HDI	3.28 [3.18, 3.39]	10.44 [10.27, 10.62]	14.34 [14.25, 14.43]	16.65 [16.52, 16.78]
LangEnglish	Mean HDI	−0.02 [−0.08, 0.03]	0.08 [−0.03, 0.19]	0.04 [−0.02, 0.10]	−0.08 [−0.17, 0.00]
LeftContextNB	Mean HDI	−0.07 [−0.11, −0.01]	−0.56 [−0.67, −0.44]	0.10 [0.05, 0.16]	−0.21 [−0.30, −0.13]
RightContextNB	Mean HDI	−0.09 [−0.16, −0.03]	−0.75 [−0.91, −0.60]	0.02 [−0.04, 0.09]	−0.17 [−0.27, −0.06]
LangEng:LeftNB	Mean HDI	0.03 [−0.05, 0.11]	0.20 [0.02, 0.37]	−0.10 [−0.18, −0.02]	0.12 [−0.01, 0.25]
LangEng: RightNB	Mean HDI	−0.01 [−0.11, 0.09]	0.21 [−0.01, 0.44]	0.01 [−0.09, 0.10]	0.00 [−0.17, 0.17]
(b)		*BW1*	*BW2*	*BW3*	*BW4*
Intercept	Mean HDI	1.28 [1.14, 1.43]	4.18 [3.87, 4.48]	3.02 [2.74, 3.30]	5.88 [5.51, 6.26]
LangEnglish	Mean HDI	−0.13 [−0.25, −0.02]	0.18 [−0.05, 0.42]	0.10 [−0.14, 0.34]	−0.08 [−0.46, 0.31]
LeftContextNB	mean HDI	0.24 [0.12, 0.37]	0.14 [−0.13, 0.42]	0.64 [0.39, 0.90]	−0.23 [−0.66, 0.20]
RightContextNB	Mean HDI	0.37 [0.22, 0.51]	0.07 [−0.29, 0.44]	0.61 [0.31, 0.91]	−0.35 [−0.87, 0.16]
LangEng:LeftNB	Mean HDI	0.15 [−0.04, 0.34]	−0.10 [−0.48, 0.29]	−0.07 [−0.49, 0.34]	0.10 [−0.52, 0.73]
LangEng: RightNB	Mean HDI	−0.20 [−0.41, 0.01]	0.11 [−0.35, 0.56]	−0.21 [−0.67, 0.25]	0.34 [−0.39, 1.07]

and L2 English, which are 321 vs 320 Hz for N1, 1305 vs 1347 Hz for N2, 2466 vs 2465 Hz for N3, and 3476 vs 3447 Hz for N4.

As expected, for most parameters there was an effect of Left and/or Right Context. For N1, N2, and N4, a non-back left or right neighbor led to a lower nasal formant than a back neighbor (see Table III); i.e., there was a heightening effect of back neighbors. The difference when compared to the overall mean was 45 − 98 Hz (0.07 − 0.75 Bark). For N3, there was only an effect of Left Context, where N3 was 47 Hz (0.1 Bark) higher for non-back than back contexts. BFs confirmed that there was evidence against H0 for most of these effects (see Table IV). For N2, N4 (left), BW1 (right), and BW3 (left), the evidence was extreme; for BW1 (left) and BW3 (right) it was strong; and for N3 (left) anecdotal. For the remaining effects, i.e., N1, N3 (right), N4 (right), BW2, and BW4, there was evidence in favor of H0.

Finally, an interaction effect of Language and Left Context contributed to the Bayesian mixed-effects model for N2 and N3 – this interaction was absent in all of the other models (see Table III). Table III shows that overall, there was a heightening effect of back Left Context on N2, where the N2 of /m/ tokens with a back left neighbor was about 86 Hz (0.6 Bark) higher than that of tokens with a non-back left neighbour. However, this difference was about 55 Hz (0.2 Bark) smaller in L2 English. For N3, there was an overall lowering effect of back Left Context, with tokens with a back left neighbor about 47 Hz (0.1 Bark) higher than those with a non-back neighbour. In L2 English, this effect was absent. Left Context thus seems to have a larger effect on L1 Dutch than L2 English. When looking at the BFs of the interaction effects, the effects disappear: for all interaction effects of Left or Right Context with Language, we found evidence in favor of H0.

The main concern of this study was whether /m/ acoustics are language-dependent within individuals. Whereas overall, we found no language effect, before concluding that /m/ remains consistent within speakers across languages, we must consider between-speaker variation in cross-linguistic differences. When looking at by-speaker means, the difference between L1 Dutch and L2 English ranged from 0.05 to 67.5 Hz (M = 13.3; SD = 12.6 Hz) for N1; from 1.5 to 180.6 Hz (M = 64.3; SD = 48.4 Hz) for N2; from 2.6 to 185.2 Hz (M = 51.5; SD = 43.1 Hz) for N3; and from 1.5 to 289.3 Hz (M = 99.7; SD = 79.3 Hz) for N4. Hence, there are speakers whose average nasal formants remain consistent across their L1 and L2, and speakers for whom the nasal formants seem to differ per language. Figure 1 shows the by-speaker means for the N1 of /m/ tokens in L1 Dutch (small circle) and L2 English (larger circle), connected by a line per speaker. Not only does the difference between the two means differ quite substantially between speakers, but also there is a between-speaker difference in which language has the highest N1. Thus, by-speaker means we confirm that there is a lack of a general pattern in cross-linguistic differences (see modelling results in Table III).

Although based on Fig. 1, there seems to be between-speaker differences in the extent of language-dependency,

TABLE IV. BFs for the different models. Values >1 indicate evidence against H0 (in italics), values <1 indicate evidence in favour of H0.

	N1	*N2*	*N3*	*N4*	*BW1*	*BW2*	*BW3*	*BW4*
LangEnglish	0.014	0.051	0.019	0.052	0.336	0.128	0.057	0.071
LeftContextNB	0.194	*>100*	*1.89*	*>100*	*25.38*	0.082	*>100*	0.123
RightContextNB	0.909	*>100*	0.009	0.705	*>100*	0.066	*84.96*	0.217
LangEng: LeftNB	0.019	0.383	0.216	0.051	0.142	0.074	0.074	0.111
LangEng:RightNB	0.018	0.239	0.011	0.014	0.255	0.085	0.118	0.185

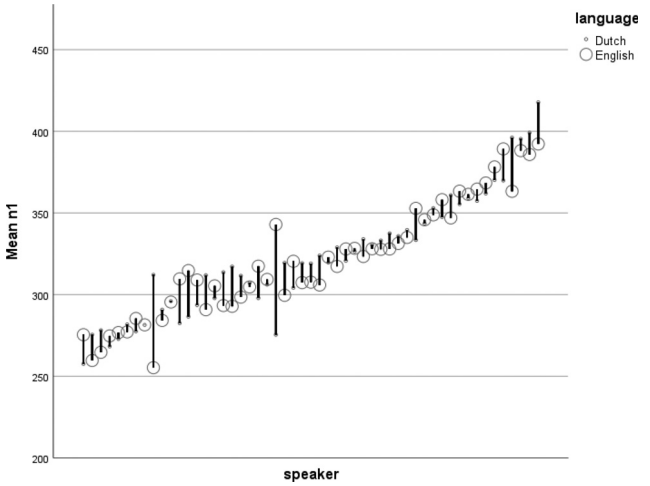

FIG. 1. By-speaker N1 means (in Hertz) for L1 Dutch (small circle) and L2 English (larger circle). Means belonging to the same speaker are connected by a line.

analysis of the SDs per speaker within either language shows that for nearly all speakers, the difference between the mean Dutch and English N1 is (much) smaller than the speaker's SD within a language. For example, the rightmost speaker in Fig. 1 (speaker ID: s065f1-1) has a between-language difference that is larger than for most other speakers, but this difference (33 Hz) is smaller than their SDs for Dutch (62 Hz) and English (44 Hz). This suggests that although for some speakers, the by-language means show larger differences than for others, the variation often still falls within the range that is expected for that speaker even within one language.

IV. DISCUSSION

This study investigated whether speakers' /m/ acoustics differ when they speak in their L2, compared to their L1. Findings showed that our group of L1 Dutch speakers had similar acoustics (i.e., nasal formants and bandwidths) in their L1 as in L2 English. This finding follows our predictions that the bilabial nasal is largely language-independent because of shared acoustic and phonological properties in both target languages in combination with the involvement of the nasal cavity: unlike the mouth, which is constantly moving during speech production, the nose is not an active articulator (e.g., Glenn and Kleiner, 1968; Rose, 2002). Indeed, some of the parameters had almost the exact same means in L1 Dutch and L2 English, and the analyses showed that there was evidence in favor of the null hypothesis for all acoustic parameters, i.e., that there were no differences between the two languages. In addition, there was evidence in favor of the null hypothesis for the interactions between language and context, confirming the absence of language-dependency in the bilabial nasal.

Whereas there was no effect of language, our findings confirm that nasal acoustics are affected by preceding and following phonetic context, i.e., coarticulation, as has been reported in prior studies (e.g., Fujimura, 1962; Kurowski and Blumstein, 1987; Smorenburg and Heeren, 2021). According to Su *et al.* (1974), /m/ acoustics are especially prone to anticipatory articulation, i.e., more affected by following than preceding context. In the current study, anticipatory articulation was not more present than carryover articulation; the preceding context affected all nasal formants, whereas an effect of the following context was absent for N3. In addition, the effect of the preceding context was confirmed more often by BFs, showing that there was evidence against the null hypothesis. Whereas Smorenburg and Heeren (2021) found that back contexts have a lowering effect on nasal formants and bandwidths, the direction of the context effects in the present study was less consistent across parameters. For most nasal formants, however, there was a heightening rather than a lowering effect of back neighbours. This discrepancy might be related to different circumstances in the two studies, e.g., speaker gender, audio quality (telephone filtering vs studio recording), and speech task. However, it could also mean that the broad categorization into back and non-back neighbours does not capture all relevant parameters for coarticulation; the distributions of possible confounds are likely to have differed between the two studies. For example, the effects of lip rounding or spreading in the back versus non-back neighbours might also affect nasal coarticulation. In addition, silent pauses (included in the non-back category) may have different acoustic effects than other non-back neighbours. Nevertheless, the presence of coarticulation effects shows that on top of anatomical properties of the nasal and pharyngeal cavity, the shape of the tongue and mouth affect the acoustics of /m/. This shows that changes in the oral configuration due to neighboring sounds can affect nasal acoustics. In the current study, any cross-linguistic differences in oral configuration, e.g., due to general articulatory settings, were less substantial than coarticulation effects. Apparently, any differences in the tension of the lips, tongue, and throat between Dutch and English (Collins and Mees, 2003) are too minimal to affect the shape of the pharyngeal cavity and the acoustics of the bilabial nasal.

For forensic speaker comparisons, our findings show that the context of /m/ may be more important when sampling tokens than the language spoken. On top of its demonstrated speaker-specificity (e.g., Glenn and Kleiner, 1968; Fujimura, 1962; Kavanagh, 2012; Rose, 2002; Schindler and Draxler, 2013; Van den Heuvel, 1996), due to its language-independence, the bilabial nasal seems a promising feature for speaker comparisons in more than one language. However, as the bilabial nasal occurs in similar positions in Dutch and English, and has similar phonetic neighbours (Collins and Mees, 2003), it is hard to generalise these findings to different language combinations which may be inherently more different in /m/ contexts. In addition, there may be larger cross-linguistic differences for speakers who are less fluent in their L2, as they may for instance use more careful articulation in their L2 than in their L1 (e.g., de Jong and Mora, 2017), thus introducing differences in coarticulatory timing or creating a difference

in the configuration of the articulators between their two languages. Hence, the bilabial nasal cannot be considered a panacea for performing cross-linguistic speaker comparisons before more research has been done on different language combinations and competency levels. For instance, when two languages differ more strongly in the phonetic contexts in which /m/ occurs, e.g., for phonemic or phonotactic reasons, this may lead to systematic differences in nasal acoustics between the languages. Moreover, a wider range of cross-language speech features would be needed to arrive at a sufficient strength of evidence to contribute useful information in casework.

More importantly, by-speaker means showed that whereas some speakers remained almost entirely consistent across their L1 and L2, other speakers showed somewhat larger differences—although for most speakers, between-language variation fell within the range of within-language variation. Thus, bearing in mind between-speaker differences in this cross-linguistic within-speaker consistency, the bilabial nasal may be a good starting point for forensic speaker comparisons where there is speech material in more than one language. When differences are found, this can then not automatically lead to the exclusion of the speaker as a possible candidate; nevertheless, that would be rare in a speaker comparison anyway.

Previous work on language-dependency of /m/ is scarce, which makes this one of the first studies to show acoustically how similar nasals in different languages can be. Previous empirical studies measuring /m/ acoustics are difficult to compare due to different speaker groups and other methodological choices. In addition, due to the speaker-specificity of nasals and anatomical differences in nasal cavities between individuals (e.g., Stevens, 1998) but also speaker groups (e.g., Ohki *et al.*, 1991), it is difficult to draw conclusions concerning cross-linguistic differences using a between-speaker approach (cf. Ng *et al.*, 2012). To our knowledge, the current study is the first to take a within-speaker approach to this question, showing that overall, /m/ acoustics may be language-independent.

In terms of language learning, our findings suggest that no language-specific /m/ pronunciation has to be taught when learning L2 English from an L1 Dutch background. Even if some speakers use slightly different realizations in L2 English, at this point, there is no reason to believe that subtle differences in /m/ acoustics affect perception or can result in misunderstandings.

V. CONCLUSIONS

Overall, our findings imply that the bilabial nasal is not language-dependent within speakers of Dutch and English and thus would be suitable to use in cross-linguistic forensic speaker comparisons. When compared to filled pause vowels (see de Boer and Heeren, 2020), /m/ remained more stable in a cross-linguistic setting and did not show any indications of language-dependency. However, caution is still recommended when sampling /m/ from different languages; even when the pronunciation of /m/ itself would not depend on the language spoken, its acoustics may be influenced by between-language differences in phoneme inventories and phonotactics, as suggested by the context effect we found.

ACKNOWLEDGMENTS

This research was supported by the Dutch Research Council (NWO VIDI Grant No. 276-75-010). We thank Beau Boogaart and Carolina Lins Machado for their assistance in making the orthographic annotations. Our thanks also go to Professor Dr. Hugo Quené and Dr. Cesko Voeten for their statistical advice.

[1]As suggested by the anonymous reviewer, besides being potentially useful in cross-language forensic speaker comparisons, language independence of certain speech segments may save the necessity of building a separate reference corpus with population data in the speaker's L2. In forensic settings, reference corpora are used to determine the typicality of a speech feature.

[2]Due to the limited bandwidth in telephone speech, N1 could not be measured in this study, as it is expected to be below 800 Hz.

[3]For two students in the sample, the order of the Dutch and English informal monologues was reversed.

[4]Contrast-coding was used to look at the overall language and context effect. Although Dutch was the speakers' L1 and non-back could be regarded as more neutral, we do not consider non-back Dutch /m/'s to be the default.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., and Wagenmakers, E. J. (**2015**). "An introduction to Bayesian hypothesis testing for management research," J. Manage. **41**, 521–543.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–48.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot. Int. **5**, 341–347.

Bradlow, A. R., Kim, M., and Blasingame, M. (**2017**). "Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate," J. Acoust. Soc. Am. **141**, 886–899.

Bürkner, P. C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Softw. **80**(1), 1–28.

Bürkner, P. C. (**2018**). "Advanced Bayesian multilevel modeling with the R Package brms," R J. **10**, 395–411.

Campbell, J. C., Nakasone, H., Cieri, C., Miller, D. J., Walker, K., Martin, A. F., and Przybocki, M. A. (**2004**). "The MMSR bilingual and crosschannel corpora for speaker recognition research and evaluation," in *Proceedings of the Odyssey 2004 The Speaker and Language Recognition Workshop*, May 31–June 4, Toledo, Spain (online).

Chang, C. (**2012**). "Rapid and multifaceted effects of second-language learning on first-language speech production," J. Phon. **40**, 249–268.

Chen, A. (**2009**). "Perception of paralinguistic intonational meaning in a second language," Lang. Learn. **59**, 367–409.

Clark, H. H., and Fox Tree, J. E. (**2002**). "Using uh and um in spontaneous speaking," Cognition **84**, 73–111.

Collins, B., and Mees, I. M. (**2003**). *The Phonetics of English and Dutch* (Brill, Leiden), Chap. 17, pp. 167–181.

Council of Europe (**2019**). "Common European framework of reference for languages: Learning, teaching, assessment (CEFR)," https://www.coe.int/en/web/common-european-frameworkreference-languages (Last viewed June 26, 2023).

de Boer, M. M., and Heeren, W. F. L. (**2020**). "Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English," J. Acoust. Soc. Am. **148**, 3612–3622.

de Boer, M., and Heeren, W. (**2021**). "Within-speaker consistency across languages: The realization of [m] in L1 Dutch and L2 English," in

Proceedings of the International Association for Forensic Phonetics and Acoustics, August 22–25, Marburg, Germany.
de Boer, M. M., Quené, H., and Heeren, W. F. L. (**2022**). "Long-term within-speaker consistency of filled pauses in native and non-native speech," JASA Express Lett. **2**, 035201.
de Bot, K. (**1992**). "A bilingual production model: Levelt's 'speaking' model adapted," Appl. Ling. **13**, 1–24.
de Jong, N. H., and Mora, J. C. (**2017**). "Does having good articulatory skills lead to more fluent speech in first and second languages?," Stud. Second Lang. Acquis. **41**, 227–239.
de Leeuw, E. (**2007**). "Hesitation markers in English, German, and Dutch," J. Germanic Ling. **19**, 85−114.
Eurostat (**2021**). "Number of foreign languages known (self-reported) by age," https://ec.europa.eu/eurostat/databrowser/view/EDAT_AES_L21/default/table?lang=en&category=educ.educ_lang.educ_lang_00.edat_aes_l2 (Last viewed January 21, 2023).
Fant, G. (**1970**). *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*, 2nd ed. (Mouton, The Hague), Chap. 2.4, pp. 139–161.
Flege, J. E., Aoyama, K., and Bohn, O. S. (**2021**). "The revised speech learning model (SLM-r) applied," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by R. Wayland (Cambridge University Press, Cambridge, UK), pp. 84–118.
Fujimura, O. (**1962**). "Analysis of nasal consonants," J. Acoust. Soc. Am. **34**, 1865–1875.
Glenn, J. W., and Kleiner, N. (**1968**). "Speaker identification based on nasal phonation," J. Acoust. Soc. Am. **43**, 368–372.
Gold, E., and French, P. (**2011**). "International practices in forensic speaker comparison," Int. J. Speech Lang. Law **18**, 293–307.
IAFPA (**2020**). "Code of practice. International Association for Forensic Phonetics and Acoustics," https://www.iafpa.net/about/code-of-practice/ (Last viewed January 21, 2023).
Jeffreys, H. (**1961**). *Theory of Probability*, 3rd ed. (Oxford University Press, New York).
Johnson, K. (**2003**). *Acoustic and Auditory Phonetics*, 2nd ed. (Blackwell, Oxford, UK).
Kavanagh, C. M. (**2012**). "New consonantal acoustic parameters for forensic speaker comparison," Ph.D. thesis, University of York, York, UK.
Kurowski, K., and Blumstein, S. E. (**1987**). "Acoustic properties for place of articulation in nasal consonants," J. Acoust. Soc. Am. **81**, 1917–1927.
Lo, J. H. (**2021**). "Issues of bilingualism in likelihood ratio-based forensic voice comparison," Ph.D. thesis, University of York, York, UK.
Lubker, J., and Gay, T. (**1982**). "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables," J. Acoust. Soc. Am. **71**, 437–448.
Ma, B., and Meng, H. (**2004**). "English-Chinese bilingual text-independent speaker verification," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 17–21, Montreal, Canada.
Makowski, D., Ben-Shachar, M., and Ludecke, D. (**2019**). "bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework," J. Open Source Software **4**, 1541–1549.
Mok, P. P., Xu, R. B., and Zuo, D. (**2015**). "Bilingual speaker identification: Chinese and English," Int. J. Speech Lang. Law **22**, 57–77.
Ng, M. L., Chen, Y., and Chan, E. Y. K. (**2012**). "Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers—A long-term average spectral analysis," J. Voice **26**, e171–e176.
Nicenboim, B., and Vasishth, S. (**2016**). "Statistical methods for linguistic research: Foundational ideas—Part II," Lang. Ling. Compass **10**, 591–613.
Ohki, M., Naito, K., and Cole, P. (**1991**). "Dimensions and resistances of the human nose: Racial differences," Laryngoscope **101**, 276–278.
Orr, R., and Quené, H. (**2017**). "D-LUCEA: Curation of the UCU accent project data," in *CLARIN in the Low Countries*, edited by J. Odijk and A. van Hessen (Ubiquity Press, London), pp. 177–190.
Quené, H., and Orr, R. (**2014**). "Long-term convergence of speech rhythm in L1 and L2 English," Social Ling. Speech Prosody **7**, 342–345.
Quené, H., Orr, R., and van Leeuwen, D. (**2017**). "Phonetic similarity of /s/ in native and second language: Individual differences in learning curves," J. Acoust. Soc. Am. **142**, EL519–EL524.
R Core Team (**2022**). "R: A language and environment for statistical computing," https://www.R-project.org/ (Last viewed 7 February 2023).
Rose, P. (**2002**). "Forensic speaker identification," in *Taylor & Francis Forensic Science Series*, edited by J. Robertson (Taylor & Francis, London), Chap. 6, pp. 125–173.
Schindler, C., and Draxler, D. (**2013**). "Using spectral moments as a speaker specific feature in nasals and fricatives," in *Proceedings of Interspeech*, August 25–29, Lyon, France, pp. 2793–2796.
Smakman, D., and De France, T. (**2014**). "The acoustics of English vowels in the speech of Dutch learners before and after pronunciation training," in *Above and Beyond the Segments. Experimental Linguistics and Phonetics*, edited by J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, and E. van Zanten (John Benjamins Publishing Company, Amsterdam), 288–301.
Smorenburg, L., and Heeren, W. (**2021**). "Acoustic and speaker variation in Dutch /n/ and /m/ as a function of phonetic context and syllabic position," J. Acoust. Soc. Am. **150**, 979–989.
Solé, M. J. (**1992**). "Phonetic and phonological processes: The case of nasalization," Lang. Speech **35**, 29–43.
Solé, M. J. (**1995**). "Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization," Lang. Speech **38**, 1–23.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, UK).
Su, L. S., Li, K. P., and Fu, K. S. (**1974**). "Identification of speakers by use of nasal coarticulation," J. Acoust. Soc. Am. **56**, 1876–1883.
Tabain, M., Butcher, A., Breen, G., and Beare, R. (**2016**). "An acoustic study of nasal consonants in three Central Australian languages," J. Acoust. Soc. Am. **139**, 890–903.
Tobias, J. V. (**1959**). "Relative occurrence of phonemes in American English," J. Acoust. Soc. Am. **31**, 631.
Van den Heuvel, H. (**1996**). "Variability in acoustic properties of Dutch phoneme realisations," Ph.D. thesis, Radboud University Nijmegen, The Netherlands.
Van der Vloed, D. L., Bouten, J. S., and Van Leeuwen, D. A. (**2014**). "NFI-FRITS: A forensic speaker recognition database and some first experiments," in *Proceedings of the Odyssey Speaker and Language Recognition Workshop 2014*, June 16–19, Joensuu, Finland, pp. 6–13.
Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.