

Language-dependency of /m/ in L1 Dutch and L2 English

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

Leiden University Centre for Linguistics, Leiden University

{m.m.de.boer, w.f.l.heeren}@hum.leidenuniv.nl

So far, most research in forensic phonetics has been performed in a monolingual context [1]. At the same time, the majority of people are multilingual [2]. Consequently, criminal cases may involve speech samples in multiple languages, sometimes even within one recording [3]. This shows the need to explore the existence of language-independent characteristics within speakers to be used in forensic speaker comparisons. Ideally, such characteristics are highly speaker-specific and are used similarly in the two languages. The current study explores language-dependency of the bilabial nasal /m/ in a group of speakers with L1 Dutch and L2 English. Prior work in a monolingual context has shown that /m/ is among the most speaker-specific segments because of the involvement of the nasal cavity [e.g. 4]. The nasal cavity is relatively rigid when compared to the oral cavity, leading to low within-speaker variability and high between-speaker variability [4]. In addition, in both Dutch and English, /m/ is a common phoneme, which is produced similarly and is used in similar phonetic contexts [5]. Hence, this study investigates whether multilingual speakers may be consistent in their production of /m/ across languages.

Method

In spontaneous monologues of 53 female speakers from D-LUCEA [6], /m/ realizations were investigated. The speakers were L1 speakers of Dutch who learned English as an L2 and had above-average L2 proficiency. They were in their 1st month of undergraduate education at an English liberal arts and science college in The Netherlands. Speakers talked for about two minutes in Dutch and then English about an informal topic of their choice. Tokens were located automatically based on the orthographic transcription and segmented manually in Praat [7]. Tokens were excluded when voiceless, creaky, <30 ms, part of the filled pause *um*, or of a different-language word. Thus, 2,972 /m/ tokens were included (Dutch: 1,681; English: 1,291).

For each token, the following measurements were taken: duration, maximum intensity (iMax), center of gravity (CoG) and its standard deviation (SD), and the first four nasal formants (N1-4) and their bandwidths (BW1-4). To see to what extent speakers' /m/ realizations were language-dependent, linear-mixed effects models were used [8], testing the fixed factor Language (Dutch, English) and random by-speaker slopes for Language. In addition, an indication of within-speaker variability was taken using SDs per speaker.

Results

Results showed that cross-linguistic differences in /m/ acoustics within the same speakers were minor (see table 1). Only for duration and N2, the best-fitting models included Language ($\chi^2(1) = 97.2$, $p < .001$; $\chi^2(1) = 9.56$, $p = .002$). Tokens in L2 English were on average 9 ms longer than those in L1 Dutch. Note, however, that we did not control for speech rate. When looking at spectral characteristics, /m/ tokens in L2 English on average had a 31 Hz higher N2 than in L1 Dutch. The other spectral measurements did not differ across the speakers' languages.

Speakers varied somewhat in the extent to which they made adaptations in the L2: for N2, iMax, CoG, and SD, random by-speaker slopes for Language were included in the best-fitting models. Whereas across speakers, the English N2 was 31 Hz higher, for some individual speakers, it was lower or more similar to the Dutch N2 (see fig. 1).

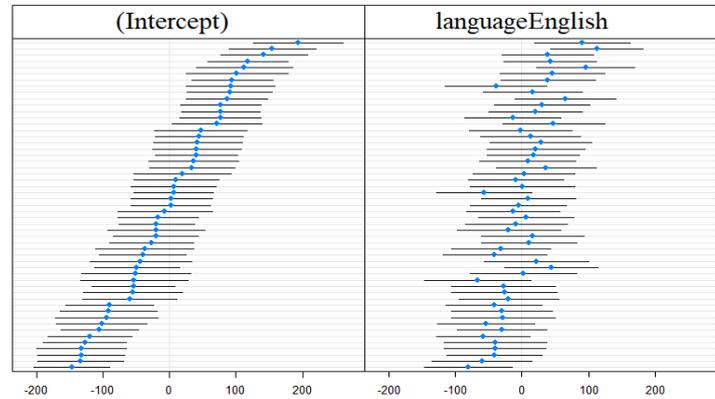
For all measurements, the means of by-speaker SDs (see table 2) were lower than the SDs across speakers (in table 1), showing that within-speaker variability seems lower than between-speaker variability. Tables 1 and 2 show that SDs were similar in both languages for most measurements, but somewhat larger in English for duration ($t(52) = -3.72$, $p < .001$).

Table 1. Overview of the means (and standard deviations) of the measurements per language.

	L1 Dutch		L2 English			L1 Dutch		L2 English	
Log10_dur (s)	-1.22	(0.15)	-1.16	(0.18)	N2 (Hz)	1,144	(272)	1,177	(303)
iMax (dB)	68	(6)	68	(6)	BW2 (Hz)	408	(352)	419	(348)
CoG (Hz)	278	(50)	277	(47)	N3 (Hz)	2,063	(378)	2,080	(368)
SD (Hz)	315	(166)	307	(166)	BW3 (Hz)	516	(379)	504	(376)
N1 (Hz)	321	(60)	322	(55)	N4 (Hz)	2,733	(332)	2,741	(325)
BW1 (Hz)	122	(66)	116	(63)	BW4 (Hz)	333	(309)	335	(328)

Table 2. Means of the by-speaker SDs in L1

	L1	L2
Log10_dur (s)	0.14	0.17
iMax (dB)	2.83	2.65
CoG (Hz)	40	38
SD (Hz)	159	161
N1 (Hz)	48	45
BW1 (Hz)	59	57
N2 (Hz)	253	269
BW2 (Hz)	328	318
N3 (Hz)	340	323
BW3 (Hz)	361	340
N4 (Hz)	304	293
BW4 (Hz)	283	305

**Fig. 1.** Caterpillar plot showing the random structure Dutch and L2 English of the N2 model, i.e. by-speaker intercepts (left) and by-speaker adaptations in the L2 (right).

Discussion and conclusion

The acoustics of /m/ seem relatively language-independent within speakers; L1 Dutch speakers showed minimal changes in their /m/ acoustics when speaking in L2 English. The feature showing the clearest cross-linguistic difference was N2, which is associated with the oral and nasal cavities [9]. Hence, despite the rigidity of the nasal cavity, some language-dependent features may remain. Based on these results, /m/ may be a useful segment for cross-linguistic forensic speaker comparisons. However, recording conditions in casework are typically worse than in the data used here, and other L2 speakers may differ in proficiency. Therefore, more research is needed to estimate the strength-of-evidence of /m/ for cross-linguistic casework, and to study /m/ in more or less advanced learners or speakers of different language combinations.

Acknowledgement: This research was supported by an NWO VIDI Grant (276-75-010).

References

- [1] Mok, P. P., Xu, R. B., & Zuo, D. (2015). Bilingual speaker identification: Chinese and English. *International Journal of Speech, Language & the Law*, 22(1).
- [2] Bhatia, T. K., & Ritchie, W. C. (2012). *The handbook of bilingualism and multilingualism*. West-Sussex, UK: Wiley-Blackwell (pp. xxi – xxiii).
- [3] Van der Vloed, D. L., Bouten, J. S., & Van Leeuwen, D. A. (2014). NFI-FRITS: A forensic speaker recognition database and some first experiments. *Proceedings of Odyssey Speaker and Language Recognition Workshop 2014*, Joensuu, Finland, June 16-19, 2014, 6-13.
- [4] Rose, P. (2002). Forensic speaker identification. In: J. Robertson (Ed.), *Taylor & Francis Forensic Science Series*. London: Taylor & Francis (pp. 125-173).
- [5] Collins, B., & Mees, I. M. (2003). *The phonetics of English and Dutch*. Leiden: Brill (pp. 167-181).
- [6] Orr, R., & Quené, H. (2017). D-LUCEA: Curation of the UCU Accent Project data. In: J. Odijk & A. van Hessen (Eds.), *CLARIN in the Low Countries*. Berkeley: Ubiquity Press (pp. 177–190).
- [7] Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [computer program]. Retrieved 3 July 2018 from <http://www.praat.org/>
- [8] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4, *Journal of Statistical Software*, 67, 1–48.
- [9] Fant, G. (1970). *Acoustic theory of speech production* (2nd ed.). The Hague: Mouton.