The effect of phonetic context on speaker information in nasal consonants

Laura Smorenburg and Willemijn Heeren
Leiden University Centre for Linguistics
b.j.l.smorenburg@hum.leidenuniv.nl, w.f.l.heeren@hum.leidenuniv.nl

Previous research indicates that linguistic context affects the speaker-dependency of speech sounds; some linguistic contexts seem to be able to convey more speaker information than others. For example, speaker classification from accented vowels is better than from unaccented vowels [5] and negative formant dynamics – associated with mouth closing gestures – show more between-speaker variation than positive dynamics [3]. However, some speech sounds are more context-dependent than others and may therefore show larger effects of linguistic context on speaker-dependency. The realisation of fricatives, for example, is highly dependent on context labialization [e.g. 2, 4]. Earlier work on two Dutch fricatives indicated that fricatives in articulatory weak and highly context-dependent positions (codas and fricatives in labialized context) were more speaker-specific than fricatives in articulatory strong and relatively context-independent positions [2]. There are also speech sounds that have been found to be relatively context-independent. The realisations of nasal consonants, for example, are less context-dependent because their resonance frequencies are largely determined by the nasal cavity. Given the inflexibility of the nasal cavity, nasals display relatively low within-speaker variability [7]. Additionally, the variability in the shapes and sizes of nasal cavities produces relatively high between-speaker variation [7]. As a result, nasal consonants have often been found to be relatively speaker-specific [e.g. 1]. This work will investigate whether the realisation of nasal consonants is dependent on phonetic context. Additionally, we investigate if speaker information in nasal consonants is context-dependent.

Although reduced relative to non-nasal speech sounds, the oral cavity does exert coarticulation effects in nasal consonants. More coarticulation effects are expected in bilabial /m/ than in alveolar /n/, because the former has no articulatory target for the tongue and is therefore subject to more context-dependent variation in tongue-position [8]. As a result, the speaker information in /m/ is expected to be more dependent on linguistic context than in /n/. Given that /n/ is articulated with the tongue in alveolar position, no effects of front phonetic context are expected for /n/. For back phonetic context, only weak coarticulation effects are expected. For /m/, we expect that larger coarticulation effects in both front and back contexts will result in higher within-speaker variation. We therefore predict lower within-speaker variation for /n/ than for /m/. However, given that timing mechanisms may lead to speaker-dependent patterns of coarticulation, there might be higher between-speaker variation in /m/.

Nasal consonants were sampled from spontaneous telephone dialogues for a set of 50 adult male speakers (Spoken Dutch Corpus: [6]). Using transcription-based forced alignment with subsequent manual correction, 2,387 /n/ onsets and 2,098 /m/ onsets and their immediate context were annotated. Neighbours segments to either side of the nasal consonant (henceforth Left and Right Context) were subsequently binary-coded for place of articulation along the front-back dimension, excluding pauses and central vowel /ə/.

Following [9], for each token, the duration, the second nasal formant (N2) and formant bandwidth (BW2) as well as the third nasal formant (N3) and formant bandwidth (BW3) were extracted over the 800-3400 Hz range. Spectral centre of gravity (CoG) and standard deviation (SD) were also extracted over the 800-3400 range. The first formant is not considered because it often merges with f_0 in nasals and because it is likely to partly fall outside of the telephone signal worked with here (300-3400 Hz).

As in [2], linear mixed-effect modelling (LMM) was used to test whether linguistic context affects nasal acoustics in spontaneous telephone speech. There are fixed factors for Left Context (front, back) and Right Context (front, back). To examine the context-dependency of speaker information, multinomial logistic regression (MLR) models are used.

In line with our predictions, preliminary LMM results show that /m/ shows larger effects of Left and Right Context than /n/: N2 shows effects of phonetic context for /m/ (Left Context: β = 21 Hz, SE = 5 Hz, t = 4.6, p<.001; Right Context: β = 63 Hz, SE = 5 Hz, t = 13.8, p<.001) but not for /n/ (Left Context: β = 9 Hz, SE = 6 Hz, t = 1.5, p = .14; Right Context: β = 12 Hz, SE = 6 Hz, t = 1.94, p = .06). Effect-sizes, however, seem relatively small. See Table 1 for means per acoustic measure per linguistic context. Preliminary MLR results furthermore indicate that relatively context-dependent /m/ has better speaker-classification accuracy (37.1%) than /n/ (31.3%).

Table 1. Means for acoustic measures from onset /m/ and /n/ over a 0.8-3.4 kHz band

	/m/					/n/				
		left context		right context			left context		right context	
measure	**total**	front	back	front	back	**total**	front	back	front	back
Dur (ms)	**68**	65	66	67	69	**63**	60	60	63	61
CoG (Hz)	**1575**	1593	1533	1628	1538	**1784**	1790	1712	1807	1753
SD (Hz)	**560**	560	553	543	570	**577**	580	580	572	584
N2 (Hz)	**1066**	1075	1047	1109	1037	**1137**	1127	1118	1146	1128
BW2 (Hz)	**108**	111	102	118	101	**170**	183	151	182	153
N3 (Hz)	**2039**	2043	2033	2035	2042	**2034**	2035	2012	2041	2029
BW3 (Hz)	**319**	316	350	296	339	**421**	405	444	405	446
N tokens	**2098**[a]	790	523	781	1216	**2387**[a]	670	625	1367	960

[a]Left phonetic context was sometimes coded as 'NA' for central vowels and for pauses, therefore, the total number of tokens is not equal to the sum of front and back left context.

Results will add to our understanding of the speaker in speech production across different linguistic contexts and speech sounds. Namely, we answer the questions whether there are locations in speech where more speaker-dependent information is available for the listener and whether this interaction differs per speech sound. The dataset will be extended to include nasal consonants in coda position and will analyse fixed factor Syllabic Position (onset, coda). A subsequent MLR analysis will indicate if the speaker-dependency of nasal consonants is dependent on linguistic context.

References

[1] Amino, K., & Arai, T. (2009). Speaker-dependent characteristics of the nasals. *Forensic Sc. Int. 185*(1–3). 21–28

[2] Anonymous. (2019).

[3] He, L., Zhang, Y., & Dellwo, V. (2019). Between-speaker variability and temporal organization of the first formant. *J. Acous. Soc. Am.. 145*(3). EL209–EL214

[4] Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward Improved Spectral Measures of /s/: Results From Adolescents. *J. Speech Lang. and Hear. Res. 56*(4). 1175

[5] McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. *Int. J. of Speech. Lang. and the Law. 13*(1). 89–125

[6] Oostdijk, N. H. J. (2000). Corpus Gesproken Nederlands. *Ned. Taalkunde 5*. 280–284

[7] Rose, P. (2002). *Forensic Speaker Identification*. Sciences New York (Vol. 20025246).

[8] Su, L., Li, K. -P., & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *J. Acous. Soc. Am. 56*(6). 1876–1883

[9] Tabain, M., Butcher, A., Breen, G., & Beare, R. (2016). An acoustic study of nasal consonants in three Central Australian languages. *J. Acous. Soc. Am. 139*(2), 890–903.