# PERCEPTION OF COARTICULATORY VOWEL NASALISATION IN GERMAN LISTENERS AND ITS RELATIONSHIP TO PRODUCTION

Esther Kunay, Philip Hoole, Jonathan Harrington

Institute for Phonetics and Speech Processing (IPS), LMU Munich
es.kunay | hoole | jmh@phonetik.uni-muenchen.de

## ABSTRACT

When making linguistic decisions, language users typically use fine phonetic details from the input they perceive. In this paper we present data from an adaptive staircase discrimination experiment in which German listeners were exposed to different patterns of nasalisation in VN sequences in a synthetical CVNCV target word. Results indicated that listeners had less difficulties in stimuli discrimination when differences affected the post-vocalic nasal consonant rather than the vowel alone, a finding which suggests that listeners are especially sensitive to those phonetic details they are experienced with. Production data from German speakers suggest that systematic differences in nasal timing occur in German depending on the nature of the postvocalic NC sequence. We further investigated whether listeners' individual perception results were related to their specific coarticulatory vowel nasalisation patterns in production.

**Keywords:** coarticulation, vowel nasalisation, adaptive staircase experiment, perception-production link.

## 1. INTRODUCTION

The emergence of distinctive nasal vowels is assumed to have its seeds in VN sequences in which coarticulatory vowel nasalisation is increased and the nasal consonant is lost over time [1, 2, 3, 4]. Specifically, vowels have been attested to become coarticulatorily nasalised more extensively if the nasal stop is followed by a voiceless obstruent, an issue that can be explained by physiological and aerodynamic principles [1, 5].

From a perceptual point of view, fine acoustic variations in vowel nasality may be used by listeners when making linguistic decisions [6, 7, 8, 9], albeit language groups differ with respect to perceptual sensitivity depending on their linguistic experience with these cues. For example, discrimination perception tests with American English and Botswana Ikalanga listeners showed that when the extent of overall nasality varied across the vowel and nasal stop, listeners could easily differentiate between two stimuli. However, discrimination became more challenging when the nasal gesture had a similar overall extent but was differently distributed across the segments [10]. This effect of perceived equivalence was stronger for the American English listeners, i.e. for the language group with more experience in coarticulatory vowel nasalisation. Further research has focused on the relationship between production patterns and perceptual sensitivity in the individual. With respect to contextual vowel nasality, findings suggest that such a relationship may be present for some language users, but not for others [9, 11, 12]. Data indicate that there is reason to postulate a greater flexibility in perception than in production, such that speakers who are inconsistent in producing a target property may as listeners nonetheless be sensitive to it (cf. [11]). This is even expected because speech perception is adaptive due to the need of constant adjustment to the linguistic environment.

In the current study we present data obtained from native speakers of Standard German, a language that does not exhibit contrastive nasal vowels. We tested whether a) German language users showed differences in perceptual sensitivity in two different tasks when the nasal gesture was variably shifted across VN segments and b) whether there was evidence for a relationship between the individual usage of coarticulatory vowel nasalisation in perception and production. Participants' production data came from a large-scale real-time magnetic resonance imaging (rt-MRI) study in which data were collected from the mid-sagittal plane to investigate the interacting gestures of the lips, tongue and soft palate during fluent speech [13, 14, 15]. In addition, synchronic acoustic data were recorded and analysed with respect to segmental boundaries. Some of the production data served as input in the current study for investigating the relationship between the usage of coarticulatory nasal cues in production and perception in the individual.

## 2. METHOD

### 2.1. Participants

An adaptive staircase perception experiment [16, 17] was run with 20 German participants (age: 19–35 years) who had previously participated as speakers in the rt-MRI study.

### 2.2. Stimuli

Stimuli patterns of AABA, ABAA, BBAB and BABB sequences were created, with the target stimulus in second or third position. As stimuli, the word [baːntə] was generated by articulatory synthesis and edited with regard to the velum lowering gesture; this was carried out with the software tool VocalTractLab [18], version 2.2 API for Windows (10 November 2017). The discrimination experiment comprised two conditions of nasalisation patterns: one in which a constant velum lowering interval was shifted along the vowel and nasal consonant, and one which contained a consistent nasal consonant but a temporally modified portion of vowel nasalisation. For both conditions, stimulus A was constructed in the same manner, representing a near-natural utterance of the target word with a nasal consonant of 80 ms and no vowel nasalisation present. Condition one (‘constant condition’) implied a constant interval of a total of 80 ms velum lowering. At the start, the difference (‘delta’) between the A and B stimuli corresponded to this maximum of 80 ms. In the course of the experiment, delta was decreased by shifting the nasalised interval as a whole in the B stimuli (details in sec. 2.3). Fig. 1 exemplifies the shifting procedure for the constant condition. Condition two (‘extended condition’) involved B stimuli with a constant nasal consonant of 80 ms but varying temporal nasalisation on the vowel, which ranged from zero to full nasalisation of 350 ms. Fig. 2 exemplifies three trial modifications of the nasalised time span during the vowel. At the beginning of the experiment, and as with the constant condition, delta corresponded to the maximum nasalisation difference between A and B (350 ms). Further, the stimuli were constructed by implementing the post-vocalic consonantal portion (/n/+/t/) with a tongue-tip gesture of a total of 160 ms. For stimulus A, the velum lowering duration was set to 80 ms, such that the nasal stop resulted from the synchronous onset of the tongue tip and the velar gesture. For all stimuli, the post-nasal voiceless stop was generated by placing a glottal gesture with the VocalTractLab specification “slightly breathy” over the remaining part of the tongue-tip gesture.

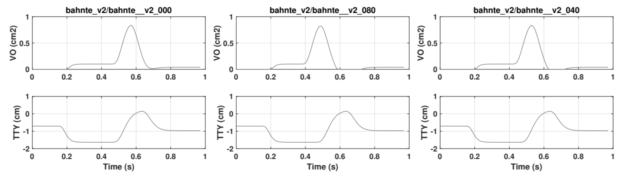


**Figure 1:** Shifting pattern for the constant condition. Left: basic configuration in stimulus A: tongue tip (TTY) and velum opening (VO) start at the same time point. Middle: VO interval shifted into the vowel by 80 ms (stimulus B): delta of stimuli A and B is at its maximum. Right: VO interval shifted into the vowel by 40 ms (stimulus B); delta of stimuli A and B is half of its maximum.

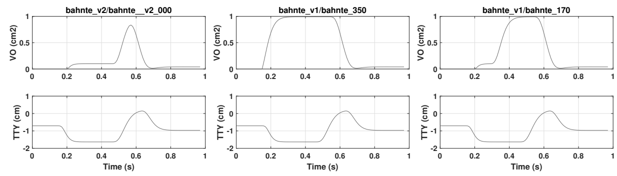


**Figure 2:** Shifting pattern for the extended condition. Left: basic configuration in stimulus A: tongue tip (TTY) and velum opening (VO) start at the same time point. Middle: overall VO of 350 ms during the vowel (stimulus B): delta of stimuli A and B is at its maximum. Right: VO of 170 ms during the vowel (stimulus B); delta of stimuli A and B is about half of its maximum.

### 2.3. Procedure

Listeners were informed that the differing stimulus was either in second or third position and they were told to press the appropriate key and to guess if necessary. Listeners were also informed that there were two conditions which would be tested separately. All participants started with the constant condition. No time limit was given and information about the outcome appeared immediately on the screen after each response was given. The overall advantage of the adaptive staircase method is an ongoing adjustment to listeners’ discrimination thresholds: depending on each response outcome, delta is either increased or decreased. In this experiment, delta was divided by 2 after two consecutive correct responses and immediately increased by 50% if one answer was wrong. Thus, the gradual delta modification was expected to approach listeners’ individual discrimination thresholds after several trials and was assumed to stay closely above and below this value. Since only delta was adapted, the trial patterns randomly alternated with respect to the direction of the nasalisation approach. As the listeners were expected to finally oscillate around their threshold boundaries, the experiment ended after 12 turning points but with a fixed upper limit of 50 trials (fig. 3). The specific by-participant thresholds were determined from a fixed number of reversals that corresponded to the six consecutive turning points with the lowest threshold values (in Fig. 3: the last six points). Thus, listeners' individual perceptual thresholds referred to a period during which the listeners were relatively constant in their decisions.

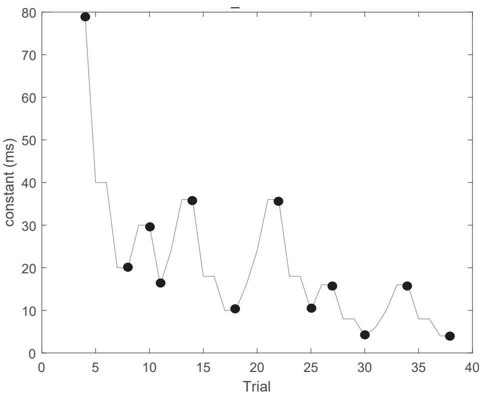


**Figure 3:** Exemplary staircase plot for one participant (constant condition). Delta is indicated by the y-axis, the trial number by the x-axis. The plot shows 12 turning points (black): one turn after the first two correct responses plus 10 turns visible in the curve plus one final turn when the twelfth turning point is reached.

## 3. RESULTS

### 3.1. Perception

Fig. 4 shows the means for the constant condition as a function of the means of the extended condition. Evidently, most participants showed overall finer discrimination abilities in the constant condition compared to the extended type. In addition, a more consistent pattern is evident for the constant condition, where 14 of the 20 data points are located below the value of 20 ms in contrast to a broader distribution for the extended condition. Moreover, responses for the two conditions did not necessarily correspond for individual listeners. For instance, while participants S14, S09 and S10 showed relatively low perception thresholds in both conditions, participant S04 had one of the lowest values in the constant condition but the highest in the extended version. Conversely, S05 exhibited a relatively high threshold in the constant

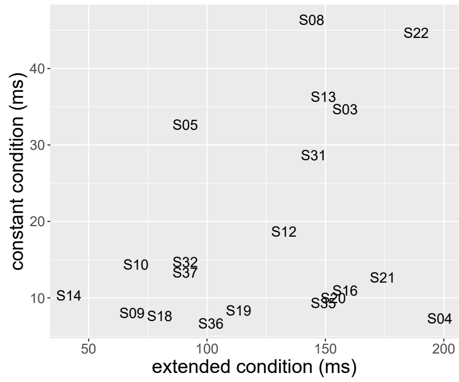


**Figure 4:** Means of the lowest six consecutive turning points for 20 listeners. X-axis: extended condition. Y-axis: constant condition.

condition but had less discrimination difficulties in the extended condition compared to most other listeners.

### 3.2. Perception and production

The perception results of the individual participants were combined with the production data obtained from the rt-MRI measurements. To elaborate a ranking in terms of by-speaker temporal vowel nasalisation in production, a total of 26 CVNV and CVNCV words in carrier phrases (CVNV n=303; CVNCV n=215) with tense vowels in pre-nasal position were analysed. The time span of the velum lowering gesture was determined by means of kinematic analyses of the velum signal obtained from the MR images. Vowel nasalisation (vow.nas) was defined as the time span between the point of maximum velocity during velum opening and the acoustic vowel offset. The overall velum lowering gesture (OVL) including the nasal stop referred to the interval between the points of maximum velocity during the velum opening and closure gesture[1]. For each participant, the amount of vowel nasalisation was divided by OVL for each target word. Thus, higher ratios corresponded to a larger extent of vowel nasalisation relative to OVL. Fig. 5 shows the by-speaker ratios ranked by the median based on the production target words. A Pearson correlation test was applied to the data of the extended condition (fig. 6 right). Results revealed no correlation between the production and perception values (r=0.230, p=0.33). For the constant condition (fig. 6 left), a Spearman correlation test[2] was run, which also did not reveal any significant correlation effect (r=-0.158, p=0.51). Data sets were combined by calculating the by-speaker ratio means, which were then defined as a function of the individual

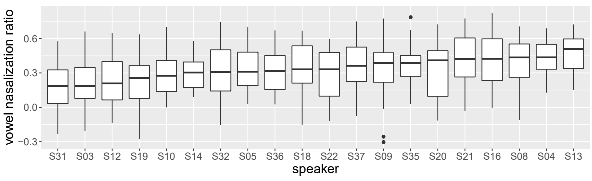


**Figure 5:** Ratio of vowel nasality to OVL (vow.nas/OVL) of target words with CVNV and CVNCV sequences for 20 speakers.

delta means from the perception test. Figure 6 shows the combined data sets for both experimental settings separately (left: constant condition, right: extended condition).

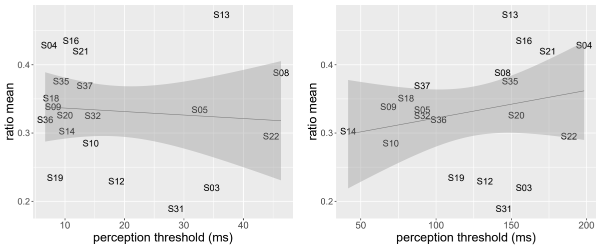


**Figure 6:** Means of the nasality ratios (vow.nas/OVL) and the perception thresholds of the lowest six consecutive turning points for 20 participants. Results are shown for the constant (left) and extended (right) condition.

## 4. DISCUSSION AND CONCLUSION

This study investigated the perceptual sensitivity of German listeners to vowel nasalisation as well as the relationship between speakers' nasality patterns in production and their discrimination skills as listeners in perception. Results revealed that listeners were overall more precise in identifying the target words when the constant interval of the nasalised portion was shifted as a whole across the vowel and nasal segment. More difficulties arose when the overall temporal extent of nasality solely varied within the vowel. Evidently, German listeners were highly insensitive to alterations of the temporal extent of vowel nasalisation and probably to the presence of coarticulatory vowel nasalisation at all. Consequently, the low thresholds in the constant condition probably do not indicate listeners' perception of vowel nasality but rather their sensitivity to other acoustic modifications coming along with the overall shifting. For example, listeners may have attended to the silent interval between the nasal offset and the offset of the /t/, which increased the more the 80 ms interval was shifted into the vowel. This increased interval may have been interpreted as a longer time span of alveolar stop closure for /t/, inducing the effect of a rather 'strong' /t/ compared to an oral stop with a short closure period similar to a weak /t/ or even /d/. As German language users are familiar with alveolar stop closure differences between /d/ vs. /t/, this might have helped them in identifying the target word more precisely. Alternatively, listeners may have relied on the stepwise absence (or in reverse trials on the increasing presence) of the nasal consonant. As with the extent of alveolar stop closure, German listeners can be assumed to have some experience with nasal weakening in certain contexts, as indicated by data discussed in [14, 15]. As shown in [15], velum lowering in CVNCV contexts was not just shifted more into the vowel compared to CVNV contexts but also reduced in its temporal extent, resulting in a shortened nasal consonant. If articulatory modifications of the nasal consonant systematically occur in production, German listeners might be sensitive to such nasal weakening patterns in perception as well, at least more than to alterations of vowel nasalisation. Considering participants' relationship between vowel nasalisation in production and perception, speakers indeed showed quite a range of anticipatory vowel nasalisation (Fig. 5). However, no correlation to their perception performances could be attested. Thus, the finding that across listeners perception skills might be related to some general experience with nasal weakening patterns in production cannot be applied to the individual language user in our data. Possibly for German, the variation patterns within a specific speaker's own production are too inconspicuous to systematically affect their perceptual sensitivity.

Overall, our experiment provides further insights into the question of which fine perceptual details listeners are able to detect when confronted with different nasality patterns. Results provide further evidence that listeners are especially sensitive to those acoustic variations they are more experienced with.

## 5. ACKNOWLEDGEMENTS

This research was funded by DFG (German Research Foundation) grant HA 3512/15-1 'Nasal coarticulation and sound change: A real-time MRI study' (J. Harrington & J. Frahm) and supported also by European Research Council Advanced Grant 742289 'Human interaction and the evolution of spoken accent' (J. Harrington).

## 6. REFERENCES


[1] P. S. Beddor, "A coarticulatory path to sound change," *Language*, vol. 85, no. 4, pp. 785–821, 2009.

[2] R. A. Krakow, P. S. Beddor, L. M. Goldstein, and C. A. Fowler, "Coarticulatory influences on the perceived height of nasal vowels," *The Journal of the Acoustical Society of America*, vol. 83, no. 3, pp. 1146–1158, 1988.

[3] J. J. Ohala and C. Jones, "The phonetics of sound change," in *Historical linguistics: Problems and perspectives*, C. Jones, Ed., 1993, pp. 237–278, longman, London.

[4] G. Zellou and M. Tamminga, "Nasal coarticulation changes over time in Philadelphia English," *Journal of Phonetics*, vol. 47, pp. 18–35, 2014.

[5] M. Ohala and J. J. Ohala, "Nasal epenthesis in Hindi," *Phonetica*, vol. 48, no. 2-4, pp. 207–220, 1991.

[6] P. S. Beddor and R. A. Krakow, "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2868–2887, 1999.

[7] P. S. Beddor, K. B. McGowan, J. E. Boland, A. W. Coetzee, and A. Brasher, "The time course of perception of coarticulation," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2350–2366, 2013.

[8] C. A. Fowler and J. M. Brown, "Perceptual parsing of acoustic consequences of velum lowering from information for vowels," *Perception & Psychophysics*, vol. 62, no. 1, pp. 21–32, 2000.

[9] G. Zellou, "Individual differences in the production of nasal coarticulation and perceptual compensation," *Journal of Phonetics*, vol. 61, pp. 13–29, 2017.

[10] P. S. Beddor, "Nasals and nasalization: The relation between segmental and coarticulatory timing," in *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 2007, pp. 249–254.

[11] ——, "The relation between language users' perception and production repertoires." in *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, U.K., 2015.

[12] M. Grosvald and D. Corina, "The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change," in *The Initiation of Sound Change: Perception, Production, and Social Factors*, M.-J. Solé and D. Recasens, Eds. Benjamins Amsterdam, the Netherlands, 2012, vol. 323, pp. 77–100.

[13] C. Carignan, P. Hoole, E. Kunay, A. Joseph, D. Voit, J. Frahm, and J. Harrington, "The phonetic basis of phonological vowel nasality: Evidence from real-time MRI velum movement in German," in *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia, 2019, pp. 413–417.

[14] C. Carignan, S. Coretta, J. Frahm, J. Harrington, P. Hoole, A. Joseph, E. Kunay, and D. Voit, "Planting the seed for sound change: Evidence from real-time MRI of velum kinematics in German," *Language*, vol. 97, no. 2, pp. 333–364, 2021.

[15] E. Kunay, "Vowel nasalization in German: a real-time MRI study," Ph.D. dissertation, Ludwig Maximilians University of Munich, Faculty of Speech and Literature Sciences, 2021.

[16] E. Gerrits and M. E. Schouten, "Categorical perception depends on the discrimination task," *Perception & Psychophysics*, vol. 66, no. 3, pp. 363–376, 2004.

[17] C. Kaernbach, "Simple adaptive testing with the weighted up-down method," *Perception & psychophysics*, vol. 49, no. 3, pp. 227–229, 1991.

[18] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PloS one*, vol. 8, no. 4, p. e60603, 2013, download: https://www.vocaltractlab.de/index.php?page=vocaltractlab-download.


---

[1] The production data indicated a clear difference with respect to the vowel nasalisation ratios relative to OVL for the two coda contexts (higher ratios in the CVNCV context). However, irrespective of the coda environment broad inspection of the relation between the production and perception data did not reveal any systematic patterns: for each perception condition, each coda condition and each vowel category (/a:, e:, i:, o:, u:, ø:, y:/), no context gave reason to consider a more detailed analysis in addition to the overall data presented in fig. 6. Although we were able to test only 20 of the former participants in the MRI production study, there was no indication that results may considerably change with a higher number of participants.

[2] The Spearman method was applied for the constant condition because data deviated from normal distribution.