PERCEPTION OF ORAL AND NASAL ALVEOLAR FLAP SOUNDS IN AMERICAN ENGLISH

Luca Garai

Department of English Linguistics, Eötvös Loránd University
garailuc@gmail.com

ABSTRACT

The perception of /t d/ alveolar flaps is well-documented in phonetic literature, unlike that of nasal flaps. The study examines how segment duration, preceding vowel quality, and vowel nasality influence the categorical perception of /t/ and /n/ uttered in a V́_ɚ flapping environment, in American English words and nonwords. A *Praat* script was used to manipulate pre-recorded words via acoustic mixing and temporal manipulation, yielding a 5-level nasality scale of preceding vowels and five durations of medial /t/ and /n/ each. A binary forced-choice perception test was conducted with six monolingual North American male participants, and a logistic regression model was used to analyze the effect of acoustic features on the perception of the medial consonant. The results show that the type of underlying consonant (t/n) and preceding vowel quality (ɪ/ɛ/æ/ʌ/ɑ) had a significant effect on consonant perception, whereas consonant duration and vowel nasality did not influence participants' decisions significantly.

Keywords: speech perception, alveolar flapping, nasalization, American English

1. INTRODUCTION

Flapping is a lenition process that results in stop consonants becoming shorter and more sonorous. In American English, it primarily targets intersonorant alveolars /t d/ [1]; for instance, the medial consonant in *shouting* /ʃaʊtɪŋ/ is shortened and voiced: [ʃaʊrɪŋ]. The most typical flapping environment is V́_V, between a stressed and an unstressed vowel [2, 3]. The main distinguishing feature between flapped and unflapped realizations of consonants is segment duration. Flapped /t d/ are usually between 10 and 40 ms long, while their unflapped realizations are articulated longer than 100 ms [4]. When a plosive is flapped, a short sonorous segment (similar to liquids) is produced instead of a longer closure+hold phase.

Many studies have investigated the perception of alveolar flaps in English [5, 6, 7, 8, 9]. Despite the detectable acoustic differences between flapped minimal pairs containing /t/ and /d/ (such as the duration of the preceding vowel due to pre-fortis clipping [10]), listeners cannot identify the underlying consonant based on these differences alone [8, 9]. Lexical bias plays a role in the perception of flap-containing minimal pairs, wherein the more frequent word form is detected more often when the minimal pairs are real lexical items [5, 7].

In recent decades, the theoretical literature documented the spread of flapping to intersonorant alveolar nasal /n/ and nasal-plosive cluster /nt/, resulting in homophony between word pairs such as winner–winter [wɪr̃ɚ] [3, 11, 12]. One previous acoustic observation regarding nasal flaps is that flaps produced from underlying /n/ and /nt/ are significantly longer in duration than oral /t/ flaps [13]. Figure 1 exhibits this comparison in sound samples recorded for the current experiment. For /n/ and /t/ appearing in a V́_ɚ environment after front and central vowels /ɛ ɪ ʌ/, the difference in consonant duration follows the previous findings of Garai [13].

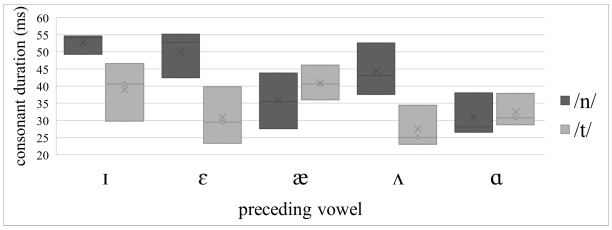

Figure 1: Duration of nasal /n/ and oral /t/ articulated in a V́_ɚ flapping environment, after stressed vowels of various vowel qualities.

During the articulation of nasals such as /n/ or /m/, the velum is lowered and air flows through the nasal cavity as well as the pharynx, in a process called velopharyngeal coupling [14]. The involvement of the nasal tract influences the acoustics of speech sounds due to the interference of resonances in the oral and nasal cavities, which

heightens the intensity in some frequency ranges and dampens it in others. Fujimura [14] observed several nasal formants (or poles) and antiformants resulting from this interference, the first of which (P0) is the most prominent around 250–300 Hz.

Nasality can spread to neighboring segments, primarily onto vowels [15]. The movement of the velum is not synchronous with the opening and closing of the lips, producing a coarticulatory effect where nasal airflow is present in the production of the adjacent vowel(s), thus changing the formant structure of these vowels [16]. Vowel nasalization is spontaneous in most languages and it can be measured by looking at the difference between the amplitude of the first oral formant (A1*) and the amplitude of the first nasal pole (P0) [17, 18]. The value of A1*–P0 is lower in nasalized vowels compared to oral ones [18]. In nonhigh vowels, P0 is identical to either the first (H1) or the second (H2) harmonic, whichever is higher in amplitude. In high vowels /i u/, the first formant is low enough that it possibly interferes with the range of P0 and in these vowels, the second nasal pole (P1) is used as a reference point for comparison [18]. Figure 2 shows the spectral difference between an oral and a nasalized realization of the same quality of vowel; the main difference lies in the relative intensity of the lower frequency range compared to the intensity of the first formant, and the first few harmonics have a higher intensity than A1 in nasalized vowels.

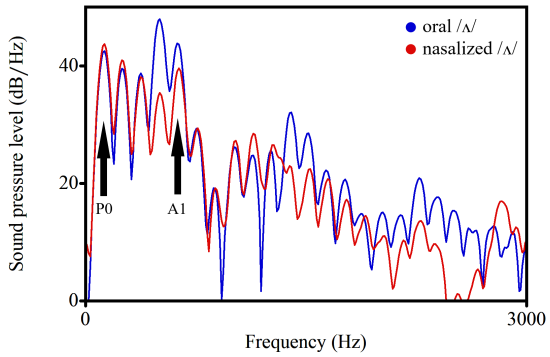

Figure 2: Spectral slice of oral and nasalized /ʌ/, with arrows pointing to the amplitude of the first nasal pole (P0) and the first oral formant (A1).

The difference between A1*–P0 values in oral and nasalized vowels is not constant throughout the duration of the vowel. Due to the coarticulatory effect of neighboring consonants, the beginning and/or the end of the vowel (depending on which side has a neighboring nasal consonant) is considerably more nasalized than the rest of the segment. This change can be observed in Figure 3 in vowels of various qualities. The A1*–P0 value of post-oral vowels changes based on whether the following consonant has a nasalizing effect on them. Vowels before nasal /n/ tend to get lower in their A1*–P0 values throughout the duration of the segment, whereas the same cannot be said for vowels before oral /t/.

This study focuses on the perception of oral and nasal alveolar flap sounds by North American listeners and aims to find preliminary cues as to whether there are distinct acoustic features (or clusters of features) by which listeners categorize the underlying consonant heard in a given sample. Based on the phonological and phonetic literature discussed above, the following two research questions are put forward:

1. How does the duration of the target consonant influence the perception of underlying /t/ and /n/ uttered in a flapping environment?
2. How does the quality and the degree of nasalization of the preceding vowel influence perception of underlying /t/ and /n/ uttered in a flapping environment?

2. METHODS

Five pairs of bisyllabic English (non)word minimal pairs were selected as the stimuli for the experiment, as seen in table 1. Each pair of words contained a stressed first vowel of a different quality, /t/ or /n/ as the medial consonant, and the initial consonant was /ʃ/ in all cases for ease of segmentability. The stimuli were read aloud three times by a 31-year-old male speaker from North Carolina and were recorded with a Blue Yeti USB microphone in a quiet environment. Recordings were segmented and annotated using *Praat* [19], and the duration of the medial consonant and the A1*–P0 value of the preceding vowel were extracted using VoiceSauce [20]. For the perception experiment, the recording with a stressed vowel exhibiting the highest A1*–P0 value in /t/-samples was chosen (out of three per word), while for samples with medial /n/, the one with the lowest A1*–P0 preceding vowel was selected.

Stimuli were manipulated using a *Praat* [19] script. The oral vowel preceding /t/ and the nasalized vowel preceding /n/ in every minimal pair was isolated from the word, matched in duration and pitch, then mixed in various ratios (100:0; 75:25; 50:50; 25:75; 0:100) creating a scale of nasality, and finally spliced back into the containing word. Furthermore, the duration of the medial consonant

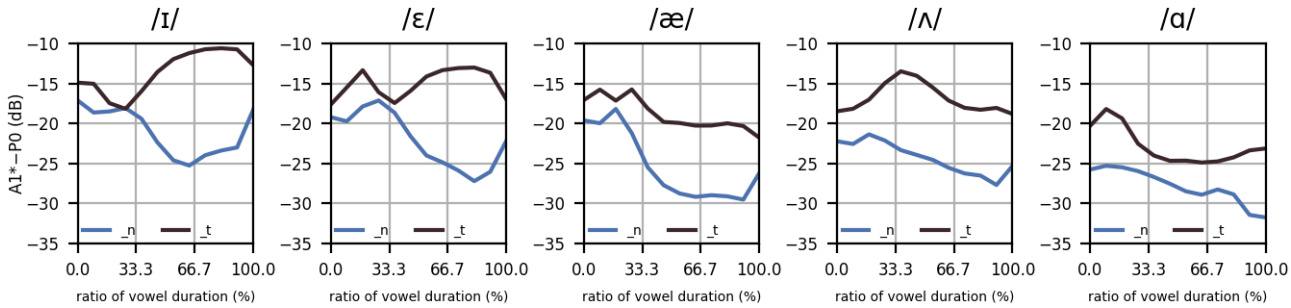

Figure 3: A1*–P0 values of vowels with various qualities articulated before /n/ and /t/ consonants throughout the duration of the vowel.

V́_	/t/	/n/
ɪ	SHITTER	SHINNER
ɛ	SHETTER	SHENNER
æ	SHATTER	SHANNER
ʌ	SHUTTER	SHUNNER
ɑ	SHOTTER	SHONNER

Table 1: Sample words recorded for the perception experiment.

was changed so that each level of vowel nasality was followed by consonants of 20, 30, 40, 50, and 60 ms, resulting in $5 * 5 * 2 = 50$ items per vowel quality, 250 target items overall.

A series of 250 bisyllabic filler items were also recorded with a word-initial /t/ or /z/, containing a diphthong as the stressed vowel, and /g k/ or /m p/ as the medial consonant, forming minimal pairs (e.g. *toager–toaker, zoomer–zooper*). Filler items were not manipulated acoustically.

Participants were recruited locally in the Raleigh, NC area with the aim of forming a homogeneous group of speakers to avoid regional and societal discrepancies. Six North American male speakers participated in the experiment with an average age of 31 years (min. 27, max. 32). All participants were native to North Carolina and were NC residents at the time of the experiment. They were all monolingual English speakers and none of them reported any conditions affecting their hearing or speech production.

A binary forced-choice test was conducted using *PsychoPy* [21] in a quiet room in each participant's home. Stimuli were presented in isolation and participants were tasked with deciding which medial consonant they heard out of the two shown on the left and right side of the screen (N and T in the case of target items), by pressing the left or right arrow key on the keyboard. The experiment contained three instances of the 250 target items each, presented in random order, with an equal number of filler items. Overall, participants judged the medial consonant in 1500 items, which were split up into nine equal parts with optional short breaks in between. Answers were recorded within *PsychoPy* [21] and exported to an Excel sheet for each participant by the software.

The analysis of the data was conducted in *RStudio* [22], with the *lme4* [23] package, using a logistic regression model. The participant's decision (N/T) for each target stimulus was used as the dependent variable, while the explanatory variables were the following: underlying consonant type and the resulting word-final nasality (n/t); duration of the consonant after manipulation (20/30/40/50/60 ms); quality of the preceding vowel (ɪ/ɛ/æ/ʌ/ɑ); nasalization of the preceding vowel after manipulation (0%/25%/50%/75%/100%). The participant was also included in the model to account for individual differences and the interaction of explanatory variables was also analyzed. McFadden's R^2 [24] was obtained to test the fit of the logistic model, and Tukey's post-hoc tests were conducted in cases where variables had a significant effect.

3. RESULTS AND DISCUSSION

Based on the results of the logistic regression model, the type of the underlying consonant (n/t) had a significant effect on the decisions of participants ($z = 2.606$; $p = 0.009$). Meanwhile, there wasn't a substantial difference in perception based on the duration of the medial consonant. The ratio of perception-based categorical decisions in relation to consonant features is shown in Figure 4 (a). Regardless of consonant duration, stimuli with an underlying /t/ were perceived as /t/ in a much higher proportion of instances (72.58% of all stimuli with underlying /t/) compared to those with

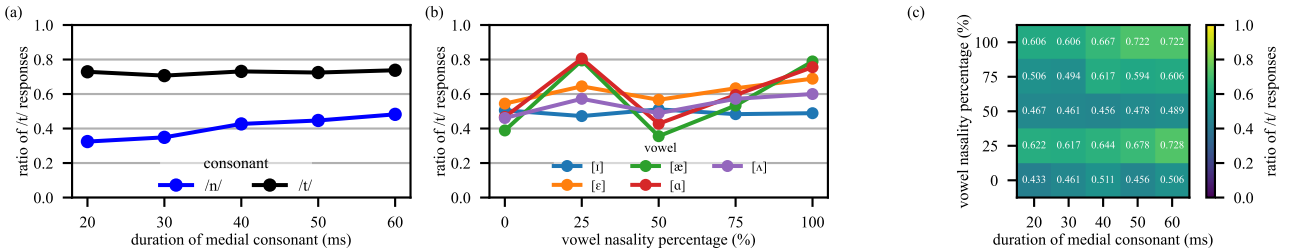

Figure 4: (a) Ratio of /t/ answers for each underlying consonant, in relation to the duration of the consonant. (b) Ratio of /t/ answers for each preceding vowel quality, in relation to the nasality of the preceding vowel. (c) Interaction of the two manipulated values (consonant duration and preceding vowel nasality).

underlying /n/ (40.58%). Whether this difference is due to the acoustic characteristics of the consonants themselves or the nasalization of the word-final [ə˞] segment supplied by the preceding /n/ cannot be determined based on the results of the current experiment. We see a positive correlation between the duration of the realized consonant and the ratio of /t/ answers in the case of stimuli with underlying /t/; this does not fall in line with previous acoustic observations detailed above in Section 1, which stipulated that /n/ uttered in a flapping environment is longer in duration than flapped /t/.

Based on the regression model, the quality of the preceding vowel had a significant effect on the perception of the medial consonant ($z = -6.535$; $p < 0.001$). The post-hoc test showed that there was a significant difference in the perception of consonants after the following pairs of vowels: /ɪ/ and /æ/ ($p = 0.005$), /ɪ/ and /ɑ/ ($p < 0.001$), /ɪ/ and /ɛ/ ($p < 0.001$), /ʌ/ and /ɛ/ ($p = 0.009$). These differences are shown in Figure 4 (b). The perception of consonant nasality before mid-high front /ɪ/ tended to differ from most other preceding vowel environments in that participants were more likely to perceive these consonants as /t/, which may be attributed to the fact that the first formant of high front vowels usually falls in the same frequency range as P0, as noted by Styler [18]. The logistic model did not show a significant effect of vowel nasality on the perception of the following consonant, therefore the degree of vowel nasalization did not influence participants' judgments in categorizing the medial consonant.

A significant interaction was observed between underlying consonant type and preceding vowel quality ($z = 6.828$; $p < 0.001$), between consonant duration and vowel quality ($z = 2.858$; $p = 0.004$), between preceding vowel nasality and vowel quality ($z = 2.526$; $p = 0.01$), and between consonant duration and preceding vowel nasality ($z = 1.976$; $p = 0.04$; this interaction is shown in Figure 4 (c)). These results reveal that while certain features do not have a significant effect on the perception of oral v. nasal alveolar consonants in a flapping environment, their influence can be detected in conjunction with other acoustic features. Furthermore, vowel quality seems to have a substantial effect on the perception of these consonants, both on its own and interacting with other features. McFadden's R^2 for the logistic regression model was $\rho^2 = 0.462$, indicating an excellent model fit, as described by McFadden [24].

4. CONCLUSION

The features with the most substantial effect on the categorical perception of /t/ and /n/ in a flapping environment were the type of underlying consonant and the quality of the preceding vowel. Underlying /t/ realizations were more likely to be perceived as /t/ than underlying /n/ realizations, regardless of consonant duration. In terms of preceding vowel quality, lexical bias might have been a potential influencing factor (in minimal pairs with real lexical items like *shitter, shatter, shutter–shunner*), since it has been shown that listeners tend to favor word forms with a higher lexical frequency [5, 7]. The formant structure of the preceding vowels might have also played a role in whether listeners identified the medial consonant as nasal.

Due to the small number of participants, the results detailed above should be viewed as preliminary. To avoid the influence of lexical bias on the results in the future, the stimulus set should be limited to nonwords or controlled for lexical frequency. Other future steps of perception research on nasal flapping are the inclusion of underlying /nt/ as the medial consonant, and diphthongs as the preceding vowel. The current results posit that there might be previously undocumented acoustic features distinguishing oral and nasal alveolar flaps.

5. REFERENCES

[1] J. S. Kenyon, *American Pronunciation. 12th ed.* Ann Arbor: George Wahr Publishing Company, 1994.

[2] D. Patterson and C. M. Connine, "Variant frequency in flap production," *Phonetica*, vol. 58, no. 4, pp. 254–275, 2001.

[3] D. Kahn, "Syllable-Based Generalizations in English Phonology," Ph.D. dissertation, Massachusetts Institute of Technology, 1976.

[4] V. W. Zue and M. Laferrière, "Acoustic study of medial /t, d/ in American English," *The Journal of the Acoustical Society of America*, vol. 66, no. 4, pp. 1039–1050, 1979.

[5] C. M. Connine, "It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition," *Psychonomic Bulletin & Review*, vol. 6, no. 11, pp. 1084–1089, 2004.

[6] N. Warner, "Cues to perception of reduced flaps," *The Journal of the Acoustical Society of America*, vol. 125, no. 3317, 2009.

[7] W. Herd, A. Jongman, and J. Sereno, "An acoustic and perceptual analysis of /t/ and /d/ flaps in American English," *Journal of Phonetics*, vol. 38, no. 4, pp. 504–516, 2010.

[8] A. Braver, "Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping," *Lingua*, vol. 152, pp. 24–44, 2014.

[9] G. Yun, "A mismatch in completeness between acoustic and perceptual neutralization in English flapping," *Korean Journal of English Language and Linguistics*, vol. 22, pp. 1133–1158, 2022.

[10] T. Scheer, "Voice-induced vowel lengthening," *Papers in Historical Phonology*, vol. 2, pp. 116–151, 2017.

[11] M. Picard, "English flapping and the feature [vibrant]," *English Language and Linguistics*, vol. 1, no. 2, pp. 285–294, 1997.

[12] B. Vaux, "Flapping in English," Linguistic Society of America, Chicago, IL, 7 January 2000, 2000.

[13] L. Garai, "Influencing factors of nasal flapping in English," in *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2021*. Budapest: Nyelvtudományi Kutatóközpont, 2021.

[14] O. Fujimura, "Analysis of nasal consonants," *The Journal of the Acoustical Society of America*, vol. 34, pp. 1865–1875, 1962.

[15] O. Fujimura and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 541–558, 1971.

[16] G. Zellou, R. Scarborough, and K. Nielsen, "Phonetic imitation of coarticulatory vowel nasalization," *The Journal of the Acoustical Society of America*, vol. 140, no. 5, pp. 3560–3575, 2016.

[17] M. Garellek, A. Ritchart, and J. Kuang, "Breathy voice during nasality: A cross-linguistic study," *Journal of Phonetics*, vol. 59, pp. 110–121, 2016.

[18] W. Styler, "On the acoustical features of vowel nasality in English and French," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2469–2482, 2017.

[19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.2.21)," http://www.praat.org/, 2022, accessed: 10-01-2022.

[20] Y.-L. Shue, P. A. Keating, C. Vicenik, and K. Yu, "VoiceSauce: A program for voice analysis," in *Proceedings of the International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 1846–1849.

[21] J. W. Peirce, J. R. Gray, S. Simpson, M. R. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. Lindelø v, "PsychoPy2: experiments in behavior made easy," *Behavior Research Method*, 2019.

[22] RStudio Team, "RStudio: Integrated development environment for R," http://www.rstudio.com/, Boston, MA, 2022.

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[24] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed. Academic Press, 1973, pp. 105–142.