TEMPORAL VS. AREA-SUM FORMULAE OF VOWEL NASALITY IN SIMULATED AND NASOMETRIC CORPORA

Michael Dow

Université de Montréal
michael.dow@umontreal.ca

ABSTRACT

Nasometric and aerodynamic methods are common in the study of vowel nasality, but their various associated “global” formulae, while correlating strongly, diverge along factors like vowel height. In this study, nasometric data for 14,000 artificial vowels were synthesized according to 14 types (nasal-oral crossover, relative slopes, etc.), and a pre-existing corpus of 4,319 French vowels was reinterpreted along the same lines. A comparison of temporal and energy-based formulae on these corpora reveals a distinct sinusoidal relationship depending on crossover point and phonological context (regressive vs. progressive). Maximal differences of 15 percentage points are noted. Energy-based values are higher for vowels with longer nasal phases, while the temporal formula provides greater values for vowels with shorter phases. These trends are robust in both corpora, except for post-nasal French /i/, with greater values according to the energy-based formula, regardless of nasal phase length.

Keywords: Vowel nasality, acoustics, simulation study, French

1. INTRODUCTION

A large variety of ways to quantify vowel nasality exists, differing not only in function of the type of data used (i.e., acoustic, aerodynamic or articulatory) but also of the nature of the research question. This study focuses on a comparison of so-called “global” expressions of vowel nasality, that is, an expression of the relative portion of a vowel which can be considered nasal or nasalized.

Two specific types of formula are considered here, dubbed *temporal* and *area-sum*. The former expresses vowel nasality as a function of the relative number of points measured (and by extension, relative duration) whose nasality exceeds an arbitrary threshold, while the latter eschews temporal information in favour of proportionate areas (nasal to total) underneath a differential curve. These formulae are most readily applicable to “split-level” data, i.e., any data source with simultaneous oral and nasal data expressed in the same units. Here we focus on acoustic data, specifically nasometry, and thus measurements of energy.

Previous studies employing the two types of measurements find that they correlate strongly but are sensitive to vowel height and phonological context (pre- vs. post-nasal) [1, 2]. However, both studies are based in real-world data from French and are thus potentially limited by language-specific constraints. In order to abstract away from these potential issues and approach the topic from a more purely mathematical point of view, this study follows up by comparing one of these corpora [2] with a simulated and more extensive corpus. This paper presents preliminary results and descriptive statistics from this study, with a focus on contextual nasalization, namely regressive (VN) and progressive (NV).

These findings have potential implications of accuracy for any empirical study quantifying vowel nasality, provided oral and nasal activity are measured separately. Additionally, these findings may further enlighten us about the complex, causal relationship between these two tracts and how it is exploited in human language.

2. LITERATURE REVIEW

Many types of instruments can used for measuring nasality, and many readings correlating with nasality can be extracted from the acoustic output (see [3] for a recent example), though not all lend themselves to the same sort of higher-level quantification. Some methods express nasality as the onset of a singular activity or correlate and/or in terms of the combined signal (see [4, 5] for extensive overviews). Meanwhile, split-level methods, which we now turn to, use the relationship between two causally-related indices of nasality and orality in tandem.

First, nasalance is a common measurement defined as the sum of nasal energy over the sum of total energy for a specific interval [6], be it single segment or an entire passage (e.g., the Rainbow Passage, [7]). While this can be used to quantify vowel nasality (e.g., [8]), it is often used clinically to assess hyponasality or hypernasality (e.g., [9]) and/or for establishing language-specific baselines (e.g., [10]). This formula, however, envisions nasality essentially as a continuum from a completely oral vowel to a completely nasal consonant. That is, any segment's nasalance can only approach 100 as the sum of its oral energy approaches 0. This somewhat complicates a direct comparison among vowels, especially given arguably inherent differences in baseline oral energy along the lines of aperture and centrality (e.g., [11]).

Nasalance can be adopted, however, on a point-by-point basis to arrive at a segment-level score of nasalization. This is exemplified by Rochet & Rochet [12] in their study of regressive nasalization in French and English. An arbitrary threshold of nasalance is used (commonly 50+) to determine whether a given point is sufficiently nasalized, and the sum of these points is taken over the total number of points to express a percentage (see also [13, 14]).

Finally, Dow [15] proposed an area-sum based formula for split-level data, called the Differential Energy Ratio (DER). Essentially, the DER calculates the proportion of nasal-dominant differential energy (nasal minus oral), defined with respect to a similar threshold, to total differential energy (cf. §3.3 for explanation).

A purported advantage of the DER is its greater granularity. The degree of difference between oral and nasal energy is directly incorporated into the final expression of nasality, as opposed to the binary transformation of the nasalance-based approach. On one hand, it is unclear to what extent, if any, this level of detail is exploited in production or perception in human speech. On the other hand, studies have shown that the otherwise significant correlation between the two types of formulae for the same dataset starts to break down for certain vowel qualities in certain phonological contexts. Dow [1] shows that French high vowels in regressive nasalization contexts are judged on average as 13% more nasal by the DER, up to approximately 25%. The cause of this is argued to be the greater susceptibility of high vowels to spontaneous nasalization, given their high intraoral pressure [1]. Meanwhile, Dow [2] finds a similar, if less pronounced, relationship on high vowels and /ø/ in post-nasal contexts.

These previous studies both used data from French and are thus necessarily limited to the vowel inventory of French, as well as its phonotactics (e.g., virtually no nasal vowels before nasal consonants) and other language-specific constraints of controlled vs. mechanical nasalization (e.g., only negligible contextual nasalization of non-high vowels). For these reasons, this study expands upon this previous work by adding a synthesized corpus.

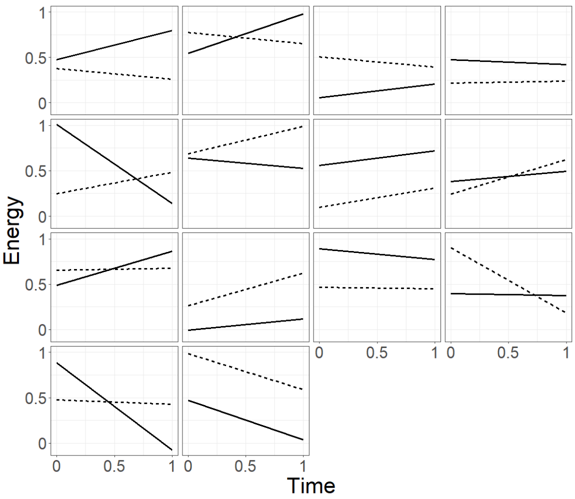

Figure 1: Illustration of all 14 vowel types. Solid line = nasal energy, dashed line = oral energy.

3. METHODOLOGY

3.1. Simulated corpus

A corpus of artificial nasometric data was first constructed in R [16]. Each vowel consisted of two series of simultaneous readings, representing oral and nasal energy, taken at equidistant intervals. A random number of intervals was chosen for each vowel between 7 and 20, essentially approximating differing durations. The initial time stamp for each vowel was a randomly generated number between 0 and 1. (Throughout, the command `runif()` was used to generate random numbers.)

Fourteen types of vowels (simplified illustration in Fig. 1) were predetermined according to every logically possible permutation of the relationships in energy (a, b) between the initial and final points within each set of readings, (c) between the initial point of each set, and (d) between the final point of each set. (Four variables with two outcomes each yields 16, with two impossible types excluded for incompatibility between slope requirements and crossover behaviour.) Essentially, this translates into the sign of the slope of each set (whether energy increases or decreases) and the intercept or lack thereof between the two sets (whether the readings of one always surpass those of the other or not).

The necessary energy relationships for each type were achieved by generating four random points between 0 and 1, sorting them, and assigning their indices as prescribed. For instance, in type 1 (first panel of Fig. 1), the initial point of the nasal readings was the third (i.e., second highest) of these points, while the final point was the fourth (i.e., highest). The initial and final oral readings of this same type were the second and first of these points, respectively.

After ensuring this relationship, the energy readings themselves were generated for each set by randomly producing n numbers between 0 and 1 (where n = predetermined intervals) between these two points. These points were then sorted for each set of energy readings according to its type (e.g., nasal increasing and oral decreasing for type 1). This was performed an arbitrary number of 1,000 times for each vowel type, yielding 14,000 artificial vowel tokens in the corpus.

3.2. Nasometric corpus

The real-world data used in this study were the same as in [15, 2]. The twenty participants (14 men, 6 women, mean age = 43.4 years) were native speakers of French from France. No significant differences were found in previous analyses along the lines of department of origin, sex or age.

Speakers read aloud a list of French expressions containing one of eleven target vowels (i.e., /a, e, ø, o, i, y, u, ɑ̃, ɛ̃, ɔ̃, œ̃/) either preceded or followed by a nasal or oral consonant, within the limits of French phonotactics. This yielded 72 total expressions. This list was read three times, each in random order, into a Glottal Enterprises hand-held nasometer (NAS-1 SEP Clinic), recorded by Praat in stereo at a sampling rate of 44.1kHz. One vowel was removed for participant error, yielding a total of 4,319 vowels.

Energy readings were extracted from each channel (corresponding to the nasal and oral microphones) at 5 ms intervals. Each channel's energy readings were then min-max normalized separately within each speaker, within each target (see [2] for reasoning and consequences of normalization).

Energy readings were then generalized. First, a linear regression was performed on each set of readings for each vowel. The intercept and slope of each regression were extracted, and the final point of each regression was estimated. Points were then randomly generated for each set (oral and nasal) and sorted according to its regression's slope. Each vowel was then classified by type (cf. §3.1). This procedure of was performed to homogenize data sources and diminish the effect of rare outliers (due to abnormally close oral-nasal readings).

Finally, the temporal (x-axis) intersection of the randomly generated oral and nasal energy readings was performed using the `line.line.intersection()` function [17] and expressed as a percentage over the total number of points for each vowel. This essentially estimates the percentage of a vowel's duration at which one sort of energy surpasses the other (if at all). This was also performed for the simulated dataset.

3.3. Formulae

Two formulae of "global" vowel nasality (i.e., the nasality of the vowel as expressed by a single number) were applied to the vowels of both corpora, namely the Differential Energy Ratio (DER) and a nasalance-based ratio along the lines of Rochet & Rochet [12]. The latter formula does not appear to have a standard name in the literature and is hereafter abbreviated NBR.

First, the DER was calculated in accordance with [15, 2]. At each point, nasal energy (y) was subtracted from oral energy (x) to provide differential energy. The absolute value of the sum of all negative differential energy was then divided by this same number plus the sum of all positive differential energy. This number was then multiplied by 100 to resemble a percentage. This is expressed by the formula in (1), where i is any measured point until the end of the vowel in question.

(1)

$$DER = 100 \times \frac{|\sum_i min(x_i - y_i, 0)|}{|\sum_i min(x_i - y_i, 0)| + \sum_i max(x_i - y_i, 0)}$$

The NBR was then calculated in the following way, as shown in (2). At each point N, the nasalance was calculated by dividing nasal energy (x) over total energy (x + y). The sum of points whose nasalance was equal to or above 0.5 was then divided by the total number of points. This number was then again multiplied by 100.

$$(2) \quad NBR = 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{x_i + y_i} \geq 0.5$$

4. RESULTS

4.1. Control vowels

We examine control vowels first, that is, vowels without intersecting oral and nasal energy lines for the simulated dataset or phonemically oral vowels in non-nasal contexts for the French data. DER and NBR scores for control vowels in the simulated dataset unsurprisingly showed perfect agreement,

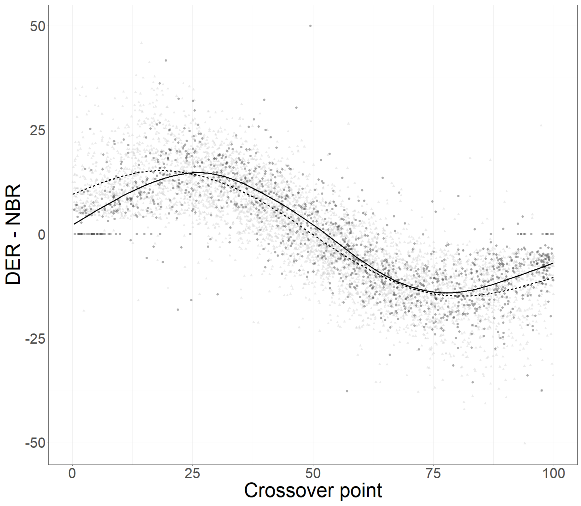

Figure 2: DER – NBR vs. crossover point. Synthesized data = darker dots, dashed line; French data = lighter triangles, solid line.

both being on average 0 or 100, depending on which type of energy dominated. On average, in the French dataset, nasal vowels have an average DER of 85.2 and NBR of 84. Oral vowels in non-nasal contexts have an average DER of 2.5 and NBR of 2.85. Essentially, both formulae agree and function as expected on these ceiling- and floor-effect tokens.

4.2. Contextual nasalization

In contextually-nasalized vowels, the NBR proved to have a linear relation with crossover point (the estimated point at which nasal energy overtakes oral), with minimal difference between corpora. This is somewhat intuitive, given the definition of the formula; the relationship, however, is imperfect, as estimated crossover points occur invariably happens between measured points, while the NBR necessarily references the first point after that crossover point. Meanwhile, the DER shows a more distinct curve towards the limits of the crossover point, such that vowels with a crossover point towards their beginning or end show floor and ceiling effects (e.g., vowels nasalized in their last quarter approach 0 DER).

Figure 2 visualizes the difference of the two (DER minus NBR) as a function of crossover point (y-axis and x-axis, respectively). In the interest of space, the crossover point of progressive nasalization is visualized with respect to the right edge, as both types of nasalization proved to be horizontal mirror images. (For instance, a vowel whose oral energy overtakes nasal energy at 25% of the vowel is given a "crossover point" of 75%.) A distinct sinusoidal relationship can be seen, such that vowels with a longer nasal phase (whether nasalized early for VN or denasalized late for NV) are judged as more nasal by the DER (max conditional means of 14.8 points for the real dataset at 26% and 15.3 at 18.5% for the synthesized dataset). Meanwhile, vowels with a shorter nasal phase (late-nasalized VN or early-denasalized NV) are judged as more nasal by the NBR to a nearly equal degree (min conditional means of -14.2 points at 78.1% for the real dataset and -14.9 at 81.2% for the synthesized dataset).

Finally, in the French data, an anomaly which bears further investigation is the behaviour of post-nasal /i/, which showed a higher difference between the DER and NBR in the 25-50% and the 50%-75% ranges than all other qualities combined. In the second quarter, /i/ tokens had on average a difference of 6.9 points, indicating the DER is higher than the NBR, versus -10.4 for all vowels combined. In the third quarter, /i/ had an average difference of 13 points, versus the 6.7 points of other qualities.

5. DISCUSSION & CONCLUSION

In this study, the DER and the NBR differed on contextually-nasalized vowels as a function of crossover point, such that higher nasality leads to higher differences in favour of the DER, and lower nasality in favour of the NBR. For instance, if nasal energy overtakes oral early within a pre-nasal vowel (VN), that vowel is likely to be judged as more nasal by the DER; the same is true for post-nasal (NV) vowels whose oral energy overtakes nasal relatively late. That this effect was found in both corpora suggests this difference is more inherent to the formulae themselves. It is surprising that the temporal component still appears to be crucial, given the attempts of the DER to divorce itself from the temporal and to model degrees of energy difference. The results of post-nasal French /i/ do suggest, however, that languages may still exploit degree of nasality (and/or velocity of change) beyond temporal properties.

In the future, changepoint detection may give us greater nuance and thus fill a gap observed in real-world data and provide a more accurate "stress test" for these formulae. Furthermore, and more crucially, perceptual research is needed to determine whether one formula more accurately reflects natural speech. Examining these questions may help us better understand vowel nasality in general, as well as providing mathematically accurate benchmarks upon which to characterize phonetic or phonological processes in language, as well as performance in clinical settings (with respect to hypo- and hypernasality, for instance).

6. REFERENCES

[1] M. Dow, “Temporal vs. area-sum measurements of vowel nasality,” 2016, Annual meeting of the Linguistic Society of America.

[2] ——, “A phonetic-phonological study of vowel height and nasal coarticulation in French,” *Journal of French Language Studies*, vol. 30, no. 3, pp. 239–274, 2020.

[3] C. Carignan, “A practical method of estimating the time-varying degree of vowel nasalization from acoustic features,” *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 911–922, 2021.

[4] R. A. Krakow and M. K. Huffman, “Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction,” in *Phonetics and Phonology, vol. 5: Nasals, Nasalization and the Velum*, M. K. Huffman and R. A. Krakow, Eds. Academic Press, 1993, pp. 3–59.

[5] V. Delvaux, *Les voyelles nasales du français : Aérodynamique, articulation, acoustique et perception*. Bruxelles: Peter Lang, 2012.

[6] S. G. Fletcher, ““Nasalance” vs. listner judgements of nasality,” *The Cleft Palate Journal*, vol. 13, pp. 31–44, 1976.

[7] G. Fairbanks, “The rainbow passage,” *Voice and articulation drillbook*, vol. 2, pp. 127–127, 1960.

[8] N. Audibert and A. Amelot, “Comparison of nasalance measurements from accelerometers and microphones and preliminary development of novel features,” in *INTERSPEECH 2011–12th Annual Conference of the International Speech Communication Association*, 2011, pp. 2825–2828.

[9] M. A. Hardin, D. Van Demark, H. L. Morris, and M. M. Payne, “Correspondence between nasalance scores and listener judgments of hypernasality and hyponasality,” *The Cleft palate-craniofacial journal*, vol. 29, no. 4, pp. 346–351, 1992.

[10] T. Sweeney, D. Sell, and M. O’Regan, “Nasalance scores for normal-speaking Irish children,” *The Cleft palate-craniofacial journal*, vol. 41, no. 2, pp. 168–174, 2004.

[11] I. Lehiste, *Suprasegmentals*. Cambridge, MA: MIT Press, 1970.

[12] A. P. Rochet and B. L. Rochet, “The effect of vowel height on patterns of assimilation nasality in French and English,” in *Pr. 12th ICPhS*, Aix-en-Provence, 1991, pp. 54–57.

[13] B. L. Rochet and F. Yanmei, “Acoustic measurements of vocalic nasality in Mandarin Chinese,” *Canadian Acoustics*, vol. 20, no. 3, pp. 53–54, 1992.

[14] A. P. Rochet and B. L. Rochet, “Patterns of assimilation nasality in English as a function of vowel height,” in *Pr. 14th ICPhS*, vol. 14, 1999, pp. 699–702.

[15] M. Dow, “Contrast and markedness among nasal(ized) vowels: A phonetic-phonological study of French and Vimeu Picard,” Ph.D. dissertation, Indiana University, 2014.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[17] D. C. Sterratt, D. Lyngholm, D. J. Willshaw, and I. D. Thompson, “Standard anatomical and visual space for the mouse retina: computational reconstruction and transformation of flattened retinae with the retistruct package,” *PLoS Computational Biology*, vol. 9, no. 2, p. e1002921, 2013.
