Discriminant analysis of nasal vs. oral vowels in French: comparison between different parametric representations

Véronique Delvaux and Alain Soquet

Laboratoire de Phonologie
Université Libre de Bruxelles
vedelvau@ulb.ac.be

Abstract

The purpose of this paper is to investigate the realization of the [nasal] contrast in French by performing an acoustic analysis of naturally spoken nasal and oral vowels, and by carrying out discriminant analysis on these data. Results consistently show that generic parametric representations allow to reliably discriminate between nasals and orals. A specific issue addressed in this paper is the relationship between phonetic and phonological nasalization in French.

1. Introduction

The acoustics of nasality in general, and the acoustics of nasal vowels in particular, is a very complex phenomenon, which has been the object of numerous studies [1][3][4][5][7][9] [10][11].

Nasality results from the acoustic coupling between the nasal cavities and the pharyngo-oral tract due to an opening of the velopharyngeal port during speech. This coupling implies that the acoustic spectrum of a nasalized vowel is more complex than the spectrum of its oral counterpart [4][11]. The effect on the spectrum depends on the degree of coupling. Moreover, the spectral characteristics of the nasal airways depend on the anatomy and differ for each individual. An asymetry between the left part and the right part of the cavities (relatively to the nasal septum) might introduce additional complexity. Finally, the effect of the sinus as extra cavities is hard to determine. They also differ in shape and volume for each individual, and their actual influence depends on the weight of nasal resonances on the final output, thus on the degree of coupling [9][11].

A vast majority of the acoustic studies on nasality has been concerned with the modelling of the phenomenon. House and Stevens [7] used an electrical analog of the oral and nasal tracts in order to investigate the effects of nasal coupling for different oral configurations. Fujimura and Lindqvist [9] confirmed these first results to a certain extent using sweep-tone measurements. More recently, several simulation experiments based on articulatory models were used to study nasality in vowels [10]. Concerning the determination of the acoustic properties of nasality, these studies result in characteristics that are not easily exploitable, either because they are very precise, arising from complex spectra and strongly dependent on the parameters of the model, or because they are very general characteristics (such as a fall down of the overall intensity) when one has to take into account data obtained for several degrees of coupling, several vowels and several subjects.

However, the most consistent effect of nasal coupling is commonly admitted to occur in the region of the first formant. Modifying the spectrum of an oral vowel in the vicinity of F1 has been proved to create the perception of nasality, although essentially for non-high vowels [3][6]. Thus, a question still remains: is this difference in the first formant region the only cue used by the listener to detect a nasal vowel or are there secondary cues, in particular for high vowels? And, in more general terms : is this information about nasality always available in the lower part of the spectrum or are there redundancies all over the spectrum?

A related issue is the actual phonetic realization of the [nasal] phonological contrast. In many languages of the world (where nasalization is distinctive), additional articulation modifications support the lowering of the velum to convey the [nasal] contrast. In French in particular, nasal vowels are quite different from their phonological oral counterparts. Young adults from Paris usually pronounce lowered /ɛ̃/, rounded /ɔ̃/, and raised /ɑ̃/ [12]. The same tendency is observed for Belgian speakers, who also usually conserve the nasal /œ̃/ in their phonological inventories. Then, what priorily signals (phonological) nasalization in French? Do the four nasal vowels still belong to a single class or are they categorized as (particular) additional units of the main set?

The purpose of this study is to investigate the realization of the [nasal] contrast in French by performing an acoustic analysis of naturally spoken nasal and oral vowels, and by carrying out discriminant analyses on this dataset. The basic assumption is to avoid directing the analysis, so the acoustic dataset is made up of several common parametric representations of speech that are computed in the middle of each vowel. The discriminant analysis is used in order to test hypotheses about the compared properties of nasal vs. oral sounds, and to address in particular the issue of the relationship between phonetic and phonological nasalization. A subset of four "nasalized" vowels are included in the corpus to this purpose.

The following questions are addressed: is it possible to discriminate between nasals and orals from generic parametric representations measured in the middle of the vowels? What does the discriminant analysis use to achieve this goal? How can it be related to the acoustic theory of nasalization? How can the performances of the oral vs. nasal and the oral vs. nasalized discriminations be compared?

2. Material and methods

2.1. Corpus and subjects

The corpus was made of a subset of NV, CV, VC, VN, and NVN items, where N=/m/, C=/s,p/, and V is one of the 4 oral vowels /a, ɛ, ɔ, œ/ and the 4 nasal vowels /ɑ̃, ɛ̃, ɔ̃, œ̃/. In addition, four [VṼV] sequences were included, for which the speakers were asked to maintain carefully the articulators in the same position throughout the sequence except for the velum. Four phonetically-trained Belgian native French speakers took part in the experiment. Their task was to repeat three times the items presented on a sheet of paper. A total of 108 oral vowels, 72 nasal vowels and 12 nasalized vowels were recorded for each subject.

2.2. Data collection

The recordings were carried out in a soundproof room by means of a microphone (Neumann, U87A i P48) placed, for each speaker, at a constant distance from the mouth. Signals were recorded on a DAT (Panasonic, SV-3700) at 44100 Hz at the resolution of 16 bits and transferred numerically on a computer for further processing.

2.3. Data analysis

The parametric representations used in this study can be divided into two groups: those based on the Fourier spectrum, and those based on the linear prediction analysis. The first group consists of the log-energy output of 12 or 24 triangular bandpass filters (EB), the mel-frequency cepstrum coefficients (MFCC) [2], and the cepstrum coefficients (CPST). The second group includes the linear predictive coefficients (LPC), the cepstrum coefficients derived from the linear prediction coefficients (LPCPST), the line spectral frequencies (LSF) [8] and the reflection coefficients (LPRC).

All parametric representations were computed on 30 ms analysis frame, pre-emphasised and tapered by a Hamming window. For the energy band coefficients, the speech signals were sampled at 22050 Hz, and the log-energy was computed for 12 or 24 filters linearly spaced on a Hertz or mel scale (respectively EB12H, EB24H, EB12M, and EB24M). For the mel-frequency cepstrum coefficients, 24 filters on a mel scale were used, and 12 coefficients computed for a sampling rate of 11025 Hz, and 24 for 22050 Hz (respectively MFCC12, and MFCC24). For all the other representations, the number of coefficients was 12 for a sampling rate of 11025 Hz and 24 for a sampling rate of 22050 (respectively CPST12, LPC12, LPCPST12, LPLSF12, LPRC12 and CPST24, LPC24, LPCPST24, LPLSF24, LPRC24).`

The signals were downsampled at 11025 Hz or 22050 Hz when necessary. Before downsampling, the signals were filtered by means of an adequate anti-aliasing filter with a cut-off frequency of 5012 Hz and 10525 Hz respectively.

The vowels of the corpus were segmented by hand with a customized iShell application (www.tribeworks.com). All parametric representations described here above were computed in the middle of each vowel.

3. Results

Classification results presented below were obtained based on a linear discriminant analysis with the coefficients of each parametric representation as independent variables. A stepwise method based on Wilk's lambda was used (F entry and F removal tresholds were respectively 3.84 and 2.71). The prior probabilities of group membership were assumed to be equal. Classification was carried out using the "jackknife" method, in which every token is successively removed from the training set, and used as a test set.

3.1. Overall performance of the different representations

Table 1 gives the correct classification scores of the discriminant analysis for the different parametric representations. Scores are given for the distinction between the oral vs. nasal vowels (720 cases) and the oral vs. nasalized vowels (480 cases).

Table 1 shows that 12 from the 16 representations perform better for the oral vs. nasal than for the oral vs. nasalized task. In a majority of cases (12/16), the 12 coefficient representations are less performant than their 24 coefficient counterparts. Surprisingly, LPLSF, a method based on an all-pole model, performs well.

Table 1: Percentage of correct classification for the different parametric representations. /O-N/ and [O-N] respectively stand for oral vs. nasal and for oral vs. nasalized classifications.

Representation	/O-N/	[O-N]	*Representation*	/O-N/	[O-N]
EB12H	79.3	81.9	EB24H	83.2	81.7
EB12M	87.9	89.8	EB24M	93.6	90.4
MFCC12	90.6	90.6	MFCC24	93.1	92.5
CPST12	79.3	80.6	CPST24	81.0	77.7
LPC12	85.6	85.2	LPC24	84.2	88.1
LPCPST12	86.0	81.3	LPCPST24	88.3	84.4
LPLSF12	90.6	86.5	LPLSF24	92.1	89.8
LPRC12	89.6	84.8	LPRC24	88.6	86.7

Figures 1 to 3 present more detailed data about the classification results of the three best representations: EBM, MFCC and LPLSF. Figure 1 plots correct classification scores for each oral/nasal pair (180 cases) and for all vowels (720 cases). Figures 2 and 3 respectively plot performances for each subject for the oral/nasal and for the oral/nasalized pairs. It can be seen that the three methods are sensitive to both vowels and subjects. The distinction /œ/ vs. /œ̃/ is systematically less accurate than the other ones. Performances are systematically worse for subject S4 (especially for the oral vs. nasalized classification). LPLSF24 appears to be more sensitive to vowels and subjects than both EB24M and MFCC24. The performances of EB24M and MFCC24 are similar.

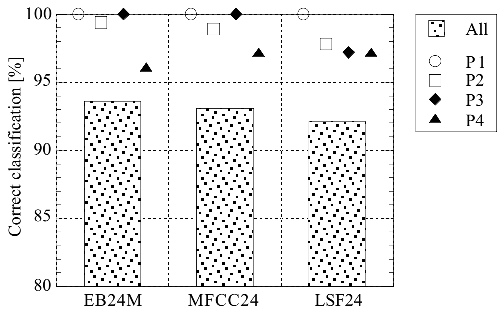

Figure 1: Percentage of oral vs. nasal correct classification for the different vowel pairs (with P1 = /a/ vs. /ɑ̃/, P2 = /ɛ/ vs. /ɛ̃/, P3 = /ɔ/ vs. /ɔ̃/, P4 = /œ/ vs. /œ̃/, and All = /a, ɛ, ɔ, œ/ vs. /ɑ̃, ɛ̃, ɔ̃, œ̃/).

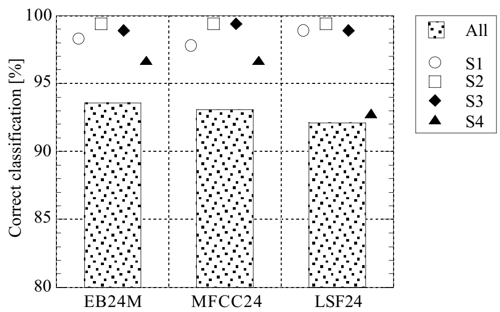

Figure 2: Percentage of oral vs. nasal correct classification for the different subjects.

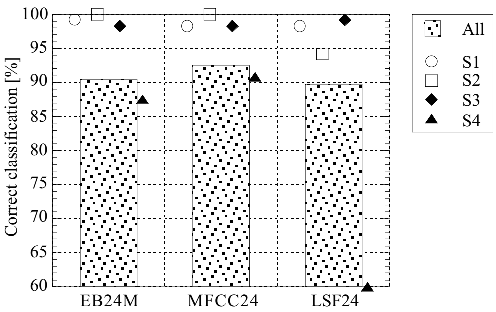

Figure 3: Percentage of oral vs. nasalized correct classification for the different subjects.

3.2. Differences between nasal and nasalized vowels

Discriminant analysis allows us to test how French vowels might be classified. Tables 2 to 4 present classification results for three different classification tasks – considering respectively all vowels as a single class, the different vowel qualities, or the nasality alone – for EB24M, MFCC24 and LPLSF24.

Tables 2 to 4 show results respectively for the subset of oral and nasal vowels (720 cases), for the subset of oral and nasalized vowels (480 cases), and for the subset of nasal and nasalized vowels (336 cases).

Table 2: Percentage of correct classification for the different parametric representations (oral and nasal vowels – 720 cases)

Representation	*8 vowels*	*4 qualities*	*Oral/Nasal*
EB24M	92.8	84.7	93.6
MFCC24	92.8	85.4	93.1
LPLSF24	89.0	86.8	92.1

Table 3: Percentage of correct classification for the different parametric representations (oral and nasalized vowels – 480 cases)

Representation	*8 vowels*	*4 qualities*	*Oral/Nasalized*
EB24M	88.3	93.3	90.4
MFCC24	87.9	92.9	92.5
LPLSF24	86.7	94.6	86.5

Table 4: Percentage of correct classification for the different parametric representations (nasal and nasalized vowels – 336 cases)

Representation	*8 vowels*	*4 qualities*	*Nasal/ Nasalized*
EB24M	83.0	89.9	83.3
MFCC24	83.3	90.2	82.4
LPLSF24	80.1	84.8	82.4

Comparing column 3 in tables 2 to 4 show that it is easier to discriminate either between oral vs. nasal or between oral vs. nasalized than between nasal vs. nasalized vowels. Thus, vowels are well discriminated along the nasal dimension, which is coherent with the presence of the [nasal] phonological contrast in French.

Besides, the difficulty to group oral and nasal vowels on the basis of vowel quality (see table 2, column 2) indicates that each nasal vowel and its phonological oral counterpart are not separated solely along the nasal dimension. They also differ in vowel quality, which allows good performances in the classification of the 8 vowels as 8 different classes.

The reverse situation is observed for nasalized vowels: they seem to match fairly well the quality of oral vowels (see table 3). This also suggests that the subjects were successful in producing the nasalized vowels with little or no other articulatory activity than the lowering of the velum, a fact which has been confirmed by informal listening.

Finally, as already noted, LPLSF24 appears to be less robust than both EB24M and MFCC24.

3.3. Interpretation of the discrimination results

The interpretation of the discrimination function itself does not allow an easy identification of the essential cues leading to the separation of the different classes. Among the representations performing well, only EB24M provides easily interpretable cues.

Table 5: Percentage of correct classification for EB24M according to the bandwidth used for the discrimination

EB24M	*Oral/Nasal [%]*	*Oral/Nasalized [%]*
1 – 8	81.0	80.2
9 – 16	77.2	80.8
17 – 24	74.3	72.5
1 – 24	93.6	90.4

Table 5 presents classification scores for the oral vs. nasal and the oral vs. nasalized tasks according to the subset of cues used in the discriminant analysis (1-8, 9-16 and 17-24). It can be seen that the performance drops when 8 successive coefficients are used instead of 24. The high frequency region (above 3800Hz) performs the worst. However, one cannot say that the low frequency region (from 0 to 1100 Hz) allows a better classification than the middle frequency region (from 1100 to 3800 Hz).

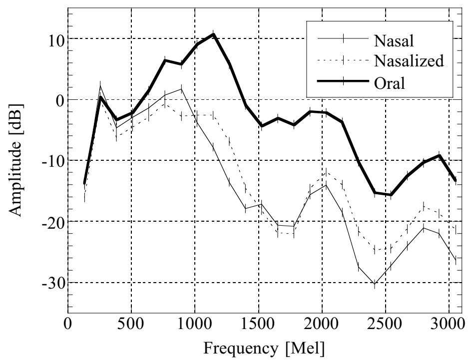

Figure 4: Mean energy band values for nasal, nasalized and oral /a/.

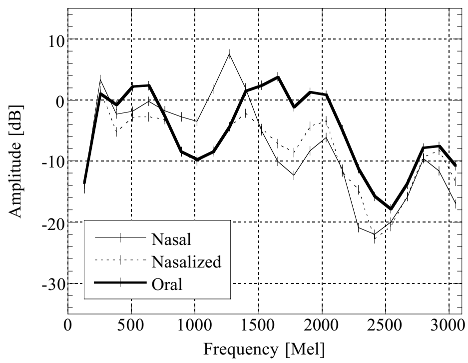

Figure 5: Mean energy band values for nasal, nasalized and oral /ɛ/.

Figures 4 to 7 show mean energy band coefficient values for each vocalic quality comparing oral, nasal and nasalized vowels across subjects. It can be seen that nasal and nasalized vowels differ from their oral counterpart throughout the frequency spectrum. The four nasalized outlines consistently follow those of their oral counterparts except for the general level of energy, which is notably inferior for nasalized vowels. The nasal vowels outlines are more different from the oral vowel outlines, especially for /ɛ/ and /œ/, which have more energy than /ɛ̃/ and /œ̃/ around the frequency of their 2nd formant.

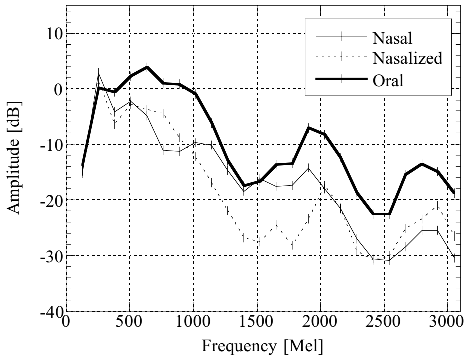

Figure 6: Mean energy band values for nasal, nasalized and oral /ɔ/.

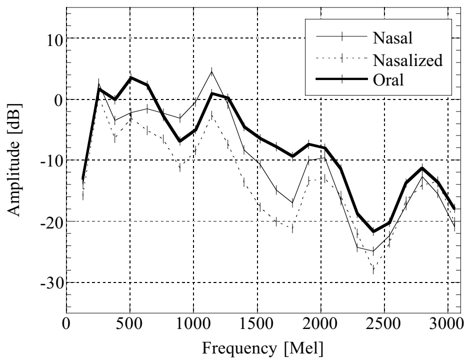

Figure 7: Mean energy band values for nasal, nasalized and oral /œ/.

4. Discussion and conclusion

Results consistently show that it is possible to reliably discriminate between nasals and orals on the basis of generic parametric representations measured in the middle of the vowels. Generally speaking, classification scores are satisfactory. They vary with the subject and with the vowel considered. It is particularly interesting that the /œ/ vs. /œ̃/ pair is the most difficult to discriminate, since the front rounded nasal vowel has completely disappeared in Parisian French and starts showing the same trend in Belgian French nowadays.

The energy band representation results clearly show that to achieve good performances more detailed information is necessary in the lower frequency range (the results in table 1 are better with a mel scale than with a Hertz scale) and that a sufficient number of bands are necessary. For the remaining representations, the weak performance improvement between the 12 and 24 coefficients shows that most of the information is located below 5000 Hz. Moreover, it has been shown that the information in mid range frequency seems to convey a similar amount of discriminative power than in the low frequency region and that both ranges convey complementary information.

A specific issue addressed in this study is the relationship between phonetic and phonological nasalization in French. Results showed that it is easier to discriminate between oral vs. nasal than between oral vs. nasalized vowels. Moreover, nasalized vowels are easily classified in the same class as their oral counterparts according to the vowel quality. This does not hold for nasal vowels. Figures 4 to 7 indicate that the nasalized and oral outlines mainly differ in their energy level, whereas nasal vowels show additional changes relatively to the frequency of the prominent resonances. The implementation of the [nasal] phonological contrast in French does not only concern the activity of the velum, but seems to rely on additional articulatory strategies and acoustic correlates.

The interpretation of the output function of the discriminant analysis does not allow a clear and decisive identification of the essential cues used to separate the different classes. We intend to use other classification techniques (like the decision trees) in a follow-up study for fine-grained interpretation of the classification strategy itself.

5. References

1. Chen, M.Y., “Acoustic correlates of English and French nasalized vowels”, ", J. Acoust. Soc. Amer., Vol. 102 (4), 1997, pages 2360-2370.
2. Davis, S. and Mermelstein, P., “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, IEEE ASSP-28, Vol. 4, 1980, pages 357-366.
3. Delattre, P., “Les attributs acoustiques de la nasalité vocalique et consonantique”, Stud. Linguist. 8 (2), 1954, pages 103-109.
4. Fant, G., “Acoustic Theory of Speech Production”, Mouton, The Hague, 1960, 328p.
5. Fujimura, O., and Lindqvist, J., "Sweep-tone measurements of vocal-tract characteristics", *J. Acoust. Soc. Amer*., Vol. 49 (2), 1971, pages 541-558.
6. Hawkins, S., and Stevens, K.N., "Acoustic and perceptual correlates of the non nasal-nasal distinction for vowels", *J. Acoust. Soc. Amer*., Vol .77 (4), 1985, pages 1560-1575.
7. House, A.S., and Stevens, K.N., “Analog studies of the nasalization of vowels”, *J.Sp.Hear.Dis*., Vol. 21 (2), 1956, pages 218-232.
8. Itakura, F. “Line spectrum representation of linear predictive coefficients of speech signals”, *J. Acoust. Soc. Amer*., Vol. 57, 1975, p S35(A).
9. Lindqvist and Sundberg, J., “Acoustic properties of the nasal tract”, *Phonetica*, Vol. 33 (3), 1976, pages 161-168.
10. Maeda, S., “Acoustics of vowel nasalization and articulatory shifts in French nasal vowels”, in Huffman, M.K. and Krakow, R.A. (Eds), Phonetics and Phonology, Vol.5, Nasals, Nasalization and the velum, 1993, pages 147-167, Academic Press, San Diego, CA.
11. Stevens, K., “Acoustic phonetics, Cambridge”, MIT Press, 1998, 602pp.
12. Walter, H., “Variétés actuelles des voyelles nasales du français”, Diachronie et variation linguistique (editor R. van Deyck), Gent, Communication et Cognition, 1994, (Studies in language), pages 223-236.

6. Acknowledgments

Véronique Delvaux is a FNRS Research Fellow. This research has been supported by the Convention ARC “Dynamique des systèmes phonologiques” 98-02, n° 226.