New Considerations for Vowel Nasalization Based on Separate Mouth-Nose Recording

Gang Feng and Cyril Kotenkoff

Institut de la Communication Parlée
INPG-Stendhal University-CNRS, ICP Campus, BP 25, 38040 Grenoble Cedex 9, France
feng@icp.inpg.fr

Abstract

In this paper an experimental setting is described which allows to separately record the speech signals emitted from the mouth and the nostrils. A series of transitions from an oral configuration to the nasopharyngeal configuration is then recorded with a speaker capable to control the movement of his velum while minimizing that of other articulators. The analysis of the recorded signals in the light of simulations with an acoustic model clearly shows that the formant evolution pattern observed in the mouth output is essentially caused by a modification of the oral tract shape due to the velum lowering and the pharyngeal constriction, while the connection with the nasal tract can be neglected in first approximation. On the other hand, the velum movement controls rather the nose output amplitude than its spectral pattern.
Index Terms: speech production, vowel nasalization, acoustic modeling, characterization, acoustic correlate

1. Introduction

It is well known that the analysis of a nasalized sound (nasal vowels, nasalized vowels or nasal consonants) is not easy. While the main articulatory gesture – the lowering of the velum that enables the connection of the nasal cavities to the oral tract – is relatively simple, the interpretation of the speech signal produced is more intricate due to the complex phenomena related to the acoustical coupling between the different cavities. It is practically impossible to find simple affiliation relations between formant frequencies and the cavities in the vocal tract, as in the case for the most of oral vowels.

There exist many studies in the literature aimed to model observed spectra by acoustic models and to find relevant acoustical and perceptual cues for a nasal sound (see [1] – [5] for example). Due to the acoustical coupling between the nasal and oral tracts, additional poles and zeros appear in the transfer function of the vocal tract. However, their positions / magnitudes depend much on the velum position, making it difficult to consider them as relevant acoustical cues. In an early study [6], we have proposed to consider the nasopharyngeal tract, realized when the velum is completely lowered, as a nasal target, and to consider the nasalization of a vowel as the transition from an oral configuration toward the nasal target.

One way to validate simulated transitions from an oral configuration to its corresponding nasopharyngeal configuration is to compare them with the spectra of real speech signals. However, when a speaker pronounces such transitions, not only the velum moves but also the other articulators. In order to reduce the differences between the simulations and the real speech, we have asked an experienced speaker to produce a series of transitions, in which he attempts to move his velum from a high position to a low position while minimizing other articulators' movements. In such a way, we expect to isolate the effect of velum lowering in the speech signal, so that the comparison with simulations is possible. An example of such transitions begins with the oral configuration of French nasal vowel [ɑ̃] (in the following, this oral configuration, which differs from the oral vowels [ɑ] and [ɔ], will be noted as [ɔ]). The speaker is asked to produce the transition from [ɔ] to the nasal vowel [ɑ̃], and then, to lower the velum until the nasopharyngeal configuration [ŋ] is reached. When comparing the simulation results with the observed speech spectra, however, it remains difficult to find a good concordance. While in the region of low frequencies (< 1500 Hz), the simulated formant trends can be roughly found in real speech spectra, the differences become more important for high frequencies.

In fact, the nasal sound signal originates from two sources: the mouth output and the nostrils' output. As it is difficult to take exactly into account the radiation conditions for each source, the total output is in general a simple sum of the two outputs in the simulations. Unfortunately, the pole-zero evolution pattern of the output varies with the relative proportion of each output changes. This adds a supplementary difficulty when comparing models with real speech signals.

One possible solution would be to obtain separately the two signals emitted from the mouth and the nostrils. In such a way, the problem of calculating the sum of the two outputs can be avoided. Moreover, the two signals can offer more information useful for modeling.

2. Experimental setting

The main difficulty to obtain mouth – nose separate recordings is to isolate the two outputs from each other. One solution using a large size insulating panel, which divides a room into two parts, is reported by [7]. Another solution consists in using a sound proof box, as reported by [8]. We have adopted the latter solution. The dimension of our box is limited to 60x60x80(h)cm. A horizontal insulating panel separates the box into two parts. Such dimensions can produce many resonance peaks in the frequency band where most speech formants are located. It is thus necessary to cover all of the inner faces of the box with alveolate sponge made for phonic isolation.

On the front face of the box, a hole is carefully cut so as to match perfectly the speaker's face. A rubber joint is attached to the edge of the hole to prevent sound leakage. In such a way the speaker's mouth and nose are all isolated. Two microphones are placed respectively in the two parts of the box, at about 10 cm from the speaker's mouth and nostrils.

In order to validate the experimental setting, a series of measurements has been carried out. First, the frequency responses of each part of the box have been measured by using a white noise excitation produced via a small loudspeaker. Although the response is not totally flat, there are no notable resonance peaks which could distort the recorded signals. The frequency response of the loudspeaker – microphone pair used for these measurements was also controlled. From these two results, the proper response of the box was obtained, which can be globally characterized as a smooth low-pass filter. This response is then used to correct the recorded speech signal.

Another important parameter of the experimental setting is the isolation between the two parts of the box. In order to measure this parameter, the lower part of the box was excited by the loudspeaker while the higher part was kept closed. The analysis of the signal captured from the microphone placed in the higher part of the box indicate that a damping of at least 20 dB is ensured for all frequencies below 8 kHz. We consider this value satisfactory for our experiment.

3. Analysis of recorded signals

The box described above was used to record several speakers. However, only one speaker (already mentioned above) could give rather stable results. The performance of this speaker had already been verified in another study using an articulograph [9]: when instructed to move only his velum, it was verified that the movement of his jaw was within a 0.03 cm range and that of his tongue (especially the tip and the middle part) within 0.2 cm. However, it cannot be completely excluded that the back part of his tongue moves backward when the velum lowers down.

Several transitions were recorded. But we limit our study to [ã] in this paper. Moreover, in order to reduce the variability, it was easier for the speaker to start systematically from a natural nasal vowel, i.e. [ã], and then either to raise the velum to reach the oral configuration - always noted as [ɔ] - or to lower the velum to reach [ŋ]. All of these transitions were produced several times and we observed a remarkable stability.

Fig. 1 and Fig. 2 show respectively the speech wave and the sonagrams for the transitions [ã] – [ŋ] and [ã] – [ɔ]. In both cases, the signals radiated by nostrils (top) and by mouth (bottom) are given.

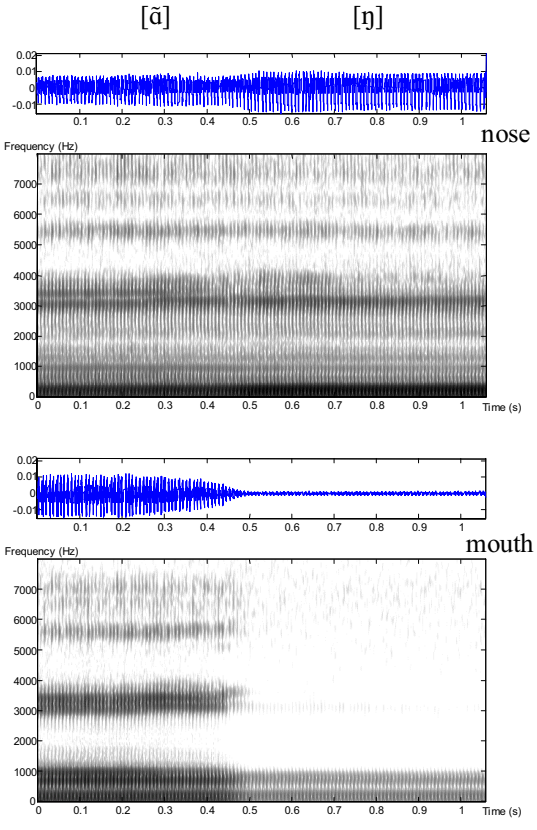

Fig. 1. Sound wave and sonagram of the transition [ã] – [ŋ]. Top: nose output; bottom: mouth output.

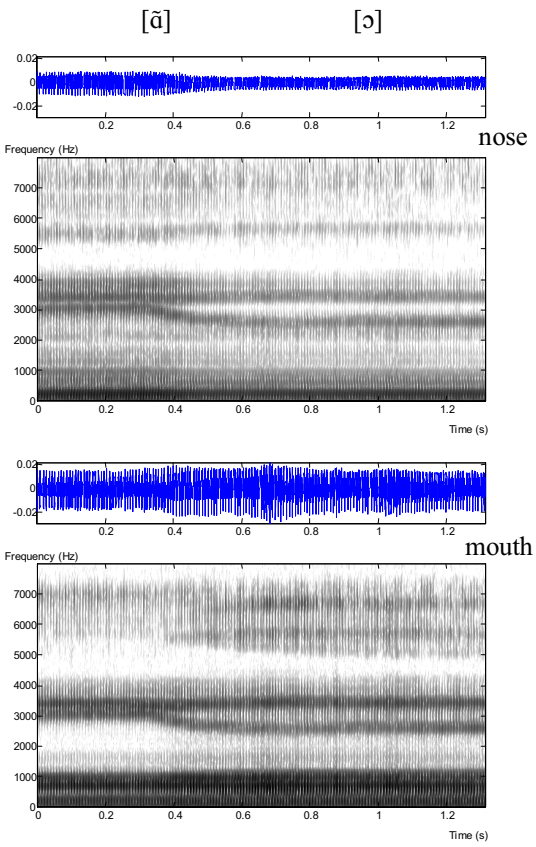

Fig. 2. Sound wave and sonagram of the transition [ã] – [ɔ]. Top: nose output; bottom: mouth output.

Several remarks can be formulated from these results. First, the strong contrast between the two outputs confirms the good isolation of the experimental setting. When the velum is completely lowered, the signal radiated from the mouth is considerably reduced (Fig. 1). The component at about 250 Hz that can be found in this signal, originates certainly from the strong emission of the same component by the nostrils in the configuration [ŋ]. The component situated at about 800 Hz corresponds to the main resonance frequency of the oral cavity. In fact, although the velum is lowered, a small part of the acoustical wave can leak and excite the oral tract.

However, we can see that in the transition [ɑ̃] – [ɔ], although the velum is raised, the nose output is not reduced to zero (Fig. 2) and remains rather strong. This is due to the difficulty of the speaker to raise completely his velum during such a transition. This can be explained by the fact that the oral part of the nasal vowel [ɑ̃] is not natural. Indeed, when the speaker produces a transition from [ɑ̃] to the true oral vowel [ɔ] without constraints on his articulatory movements, the nose output becomes much weaker.

Consider now the two transitions linked together, i.e. the evolution from [ɔ] to [ŋ]. We observe in the mouth output, that the formant F3 of [ɔ], around 3000 Hz, comes close to the formant F4 when the velum lowers, as often observed for the nasal vowels [ɑ̃] and [ɔ̃]. The same evolution of these formants can also be observed in the nose output. The question now is: having the two outputs separated, can we find a satisfactory explanation for this formant evolution pattern?

4. New considerations for acoustical modeling

The acoustical model used in this study is a standard transmission-line (with loss) model. The velum movement is simulated by a progressive decrease of the areas in the velar region of the vocal tract where the nasal tract is connected, and also by a progressive increase of the areas at the beginning of the nasal tract (see [6] for more details). Fig. 3 shows the results of such a simulation for the nasalization of [ɔ], the oral configuration corresponding to the nasal vowel [ɑ̃]. The nose output and the mouth output are separately given so that it is possible to compare them with the sonagram of the recorded signals using the experimental setting. It can be seen that, in the simulated mouth output, the F3 frequency of [ɔ] decreases when the velum lowers, while in the sonagram, this frequency increases.

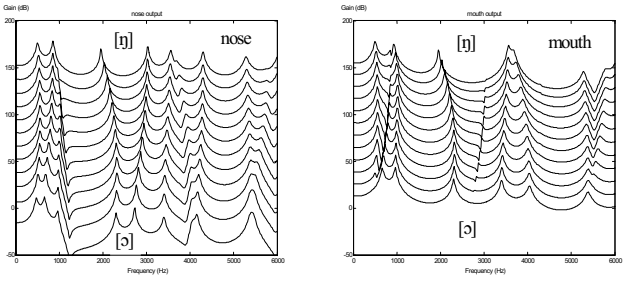

Fig. 3. Simulation of oral and nasopharyngeal transfer function variations for a velum transition from [ɔ] to [ŋ].

Although in such a comparison, the precision of the formant frequencies is not a priority, because of the approximation of the model, we think that the model should reproduce the main trends of the formant evolution.

4.1. Mouth output

What determines the evolution trends of the formants? We know that the connection of the nasal tract to the oral tract introduces several pole-zero pairs in the mouth output, modifying the initial formant structure. In order to see the effect of the nasal tract connection, the above simulation has been modified so that when calculating the mouth output, only the modification of the oral tract due to the velum movement is taken into account, the acoustic coupling with the nasal tract being eliminated. The results are given in Fig. 4. It can be clearly seen that the evolution of the F3 remains the same whether the nasal tract is connected or not. Comparing these figures, it is not difficult to observe that the formant evolution pattern depends much more on the modification of the oral tract shape, here due to the movement of the velum, than on the connection of the nasal tract.

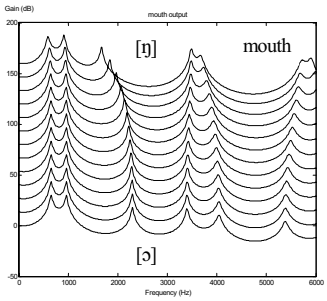

Fig. 4. Simulation of oral transfer function variations for a velum transition from [ɔ] to [ŋ]. The acoustic coupling to the nasal tract is not taken into account.

Since the above simulation does not correspond to the observed evolution for F3, we think that the modification of the oral tract should not be only limited to the movement of the velum. When producing such transitions, the speaker controls the movement of his velum, but at the same time, the other articulators may slightly move (without being noticed). We know that the tongue moves back to a greater extend for the nasal vowels such as [ɑ̃] and [ɔ̃] than for their corresponding oral vowels (see for example [10]), resulting in a constriction in the pharyngeal part of the vocal tract. It is thus not impossible that our speaker produces the same phenomenon. On the other hand, after many attempts, we could not obtain the good evolution of F3 in the simulation without introducing a constriction in the pharyngeal part of the vocal tract. Fig. 5 shows a simulation result for the mouth output (nasal tract connection neglected), obtained with a progressive constriction of the pharynx, in addition to the velum movement. The F3 evolution is now consistent with the observations.

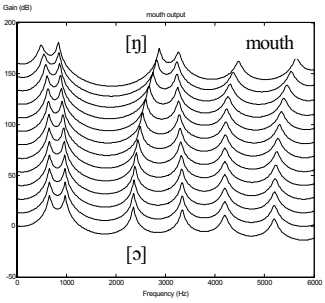

Fig.5. Simulation of oral transfer function variations for a velum transition from [ɔ] to [ŋ]. The pharyngeal constriction is taken into account.

The analysis and the acoustical modeling of the separately recorded mouth output give us a very important information: the formant evolution of this output can be fairly well explained by the sole modification of the vocal tract shape (velum lowering with the movement of other articulators). This offers a better way to understand the behavior of the mouth output in a very simple manner, since we only need to deal with a single tract without the complex coupling problem.

4.2. Nose output

The signal radiated by the nostrils shows less contrast in the sonagram with respect to the mouth output. Indeed, if we listen to these signals, we almost always hear [ŋ]. This can be explained by the fact that the nose output is strongly characterized by its spectral peak at about 250 Hz, which has a more important amplitude than the other peaks; moreover this peak varies little when the velum lowers. Rather, the velum position controls the global amplitude of the nose output.

These observations support our early hypothesis [6] : it is the presence of the nasal target characterized by a nasopharyngeal configuration ([ŋ]-type) that allows a sound to be perceived as nasal. The relative amplitude of this signal with respect to the mouth output modifies the degree of perceived nasality.

Although the acoustical modeling of the nasopharyngeal tract is theoretically very simple, the real transfer function of this tract is much more complex, due to the high complexity of the anatomic structure of the nasal tract. This complexity can also be seen in the sonagram of the nose output.

5. Conclusions

To summarize, separate mouth-nose recordings allow us to better understand the mechanism of the production for the two outputs. It has been shown that the formant evolution pattern observed in the mouth output originates mainly from the modification of the oral tract shape during nasalization, while the acoustical coupling due to the connection of the nasal tract plays only a secondary role. Moreover, the velum movement has little effect on the spectral shape of the nose output. Rather, it controls its global amplitude, modifying the degree of perceived nasality.

We propose to consider the nasal sound signal as the simple sum of these two outputs. One carries the oral tract shape message, useful to distinguish different sounds that are nasalized, while the other one signals the presence of the nasopharyngeal tract, by which the sound is perceived as nasal.

This analysis approach should not be considered as a rough simplification or as a technical method just to avoid the complex acoustical coupling problem, but as a way to find out the main relevant features in relation with the articulatory movements and to identify what can be neglected at first time.

Further studies will be mainly focused on the following directions: obtaining more mouth-nose separate acoustic data, and improving the acoustic modeling for the nasopharyngeal tract.

Acknowledgements

We would like to thank our speaker, Pierre Badin; Alain Arnal for his help to make the box and Xavier Pelorson for his help in acoustics.

References

[1] Fant, G., *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960.

[2] Fujimura, O. and Lindqvist, J., "Sweep-tone measurements of vocal tract characteristics", *J. Acoust. Soc. Am.* 19, 541-558, 1971.

[3] Maeda, S., "Acoustic correlates of vowel nasalization: A simulation study", *J. Acoust. Soc. Am*. Suppl. 1. 72, S102, 1982.

[4] Hawkins, S. and Stevens, K.N., "Acoustic and perceptual correlates of the non-nasal – nasal distinction for vowels", *J. Acoust. Soc. Am*. 77, 1560-1575, 1985.

[5] Dang, J., Honda, K. and Suzuki, H., "Morphological and acoustical analysis of the nasal and the paranasal cavities", *J. Acoust. Soc. Am*. 96, 2088-2100, 1994.

[6] Feng, G. and Castelli, E., "Some acoustic features of nasal and nasalized vowels : A target for vowel nasalization", *J. Acoust. Soc. Am*. 99, 3694-3706, 1996.

[7] Schnell, K. and Lacroix, A., "Generation of nasalized speech sounds based on branched tube models obtained from separate mouth and nose outputs", *Proc. ICASSP*, 2003.

[8] Suzuki, H., Nakai, T., Dang, J. and Lu, C.X., "Speech production model involving subglottal structure and oral-nasal coupling through closed velum", *Proc. ICSLP*, 1, 437-440, 1990.

[9] Rossato, S., *Du son au geste, inversion de la parole : le cas des voyelles nasales*, PhD thesis, INPG, France, 2000.

[10] Zerling, J.P., "Phénomènes de nasalité et de nasalisation vocalique : Etude cinéradiographique pour deux locuteurs", *Travaux Inst. Phonet. Strasbourg*, 16, 241-266, 1984.