RECOGNITION OF FRENCH NASAL VOWELS
Li Wu*, J.P. Haton*, F. Lonchamp*.

ABSTRACT

The difficulties of nasal vowel recognition are twofold. The additional zeros as well as poles in the acoustical speech, introduced by the connection of the nasal branch to the oral branch, make it impossible to acquire a stable and clear spectrum. On the other hand, analytical models are insufficient for explaining the zeros and poles appearing in the spectrum.

In this article, we shall discuss the general characteristics of French nasal vowels, and then introduce a pole-zero model in the frequency region selected by local optimization of the spectral matching. Lastly, we combine the LP all-pole model with this local optimized zero-pole model for the recognition of a database (325 words) of natural French nasal vowels pronounced by thirteen male speakers.

In the experiment, we have not chosen any reference, because we want to find speaker independent parameters. The result of the recognition shows that the combination of LP model with pole-zero model is able to help us to reveal the parametric changes of the nasal vowel /ɛ/ and /æ/. As to the /ɛ/ /ã/, we were unable to find more information about nasality.

INTRODUCTION

Nasal vowels are an important group of phonemes. Many people have been studying systematically their features since the 50's, and have revealed many of their external characteristics. The pronunciation of a nasal vowel is different from that of an oral vowel. A nasal vowel is produced by introducing acoustic coupling between the oral and nasal cavities at a point which is about halfway along the vocal tract between the glottis and the lips (ref 1). So, beside of the three principal formants reflecting the oral resonator, some additional zeros and poles are introduced in the nasal signal by the coupling. These additional zeros and poles make the nasal signal so complicated that we encounter difficulties in LPC analysis, since LPC is a powerful tool for the analysis of a signal characterized by an all-poles model. In the following section, we analyse the four French nasal vowels with LPC and zero-poles model.

PROPERTIES OF NASAL VOWELS

Soft palate is a valve that controls the connection of the nasal cavity to the oral cavity, the opening of the soft palate makes speech nasalized, else oral. Here we do not distinguish the nasality in the sense of articulation from perception. The speech signal is considered as nasal as long as the nasal cavity is connected to the oral cavity. The sound pronounced after connection is the result of the interaction of three resonators: oral, nasal cavities and velopharyngeal near glottis. These three cavities form their own formants in output signal. These interaction makes analysis very difficult:

(a) By comparison with the oral vowel, the nasal speech spectrum no longer possess a stable form. The relative position of formants will be perturbed. The most evident fact is the appearance of additional formants and energy redistribution.

(b) The level of nasality depends upon the soft palate opening. For this reason, we say that nasality depends on the speaker and on the context. Sometimes the nasalization takes place at the initial stage of a vowel, and sometimes at the last stage.

*CRIN-INRIA : B.P 239, 54506 Vandœuvre-les-Nancy cedex, France
*Institut de Phonétique de Nancy : B.P 3397, 54015 Nancy cedex, France
(c) Tools for the analysis have not yet been perfected. The classical LPC is not suited to the signal with zeros in their spectrum. The ARMA model is applicable in the case where the signal contains poles as well as zeros. However, the choice of the polynomial order is not easy to decide.

According to the theory proposed by Fant and developed by Fujimura, M.Mrayati (ref 2) simulated on computer the mutual influence of oral formants and nasal formants, and described the possible spectrum form of a nasal vowel.

F.Lonchamp (ref 3) has taken a detailed analysis for natural nasal vowels by means of the observation on their DFT spectrum. He has given some qualitative conclusions for four nasal vowels. We define F'1 F'2 F'3 as the frequencies corresponding to oral formants in the nasal case, and Fin and Ain as the ith nasal formants or ith nasal antiformant.

/ɛ/ is characterized by one peak within 500 and 800 Hz, composed of two formants F'1 and F'2 and one zero Ain.

/æ/ is characterized by the first F'1 near 700Hz and the second formant F'2 near 1000Hz.

/œ/ is characterized usually by first formant about 600 - 700 Hz and second formant which is relatively intense in the region near 1600 Hz.

/œ/ is characterized by the first formant about 700 Hz and the second formant about 1250 Hz.

/œ/ and /œ/ are very similar in their formant frequencies. In reality, there is a tendency in French that people do not distinguish them in articulation.

HOW TO RECOGNIZE NASAL VOWEL

Nasal articulation is evidently different from that of oral, but the process of perception is completed in the same auditory system. The fact that an auditor can distinguish nasal vowels without any difficulty reveal that the information about nasal perception must be included in the frequency range easily perceptible by human ear — i.e. in formants.

In order to measure precisely the formant frequencies, we have chosen a zero-pole model, which has the following form:

$$H(z) = \frac{\sum_{i=0}^{M} b_i z^{-i}}{\sum_{i=0}^{N} a_i z^{-i}}$$

where H(z) is the system transformation function, M and N are respectively the orders of numerator and denominator polynomial. To avoid the difficulties in choosing these values, we have limited the analysis range of the model within a region where the number of zeros and poles could be roughly predicted by the past experience. Choosing the order of M and N according to the number of zeros and poles that may appear makes the model locally optimized for that region. The analysis error introduced by incorrectly chosen orders could be reduced. We use LPC method to analyse formants of the oral and the zero-pole model above to analyse nasal vowel. Figure 1 shows F1 F2 and F3 plans of formants, pronounced by a french male speaker. Phrases are the story of "La bise et le soleil..." (the wind and the sun). We can see from this figure that the distributions of F1 F2 and F3 of a nasal vowel is different from the ones of the corresponding oral vowel. Formants actually yield some information about nasality.

Figure 2 describes spectrographic evolution of four nasal vowels. It is clear that for front vowels /ɛ/ /œ/, the first formant remains in a higher frequency range or move toward the high frequency range. We have noticed that it is the spectrum between F'2 and F'3 that is most disturbed. For back vowels /ɔ/ and open /a/, the most evident characteristic is that the F'1 and F'2 are closed to each other. For /ɔ/, F'1 and F'2 usually become one single peak with a
wide band.

The four French nasal vowels are generally characterized by high frequencies of F'1 and F'3 in comparison with oral vowels. So, if we can extract more precisely the formant frequencies F'1, F'2, F'3, we should be able to identify nasal by their relative formant positions.

**Identification of Nasal Vowels**

According to the rough study on nasals, we have focused our observation on a corpus of thirteen male pronunciations, where each speaker repeats two times thirteen isolated carefully chosen words. With a 25.6 ms Hamming window and 0.9 preaccentuation, we have analysed this corpus by classical LPC and by the zero-poles model mentioned above, with the frequency range of analysis limited within 0-3200 Hz.

We use a low order LPC to calculate energy in order to automatically fix the analysis window position. In general, maximum value of the energy curve does not represent a real vocalic kernel, but only one sign for a vocalic region. Consequently we try to find another sign from the spectrogram whose stability offers a second indication for that region. We fix finally the analysis window position by combining the indication from the energy and from the spectrogram.

In order to overcome the instability of parameters which is caused by unmatched model, LPC analysis is carried out four times around the fixed window position, and we take the average of formants and bands. We then match properly the formants of the zero-poles model with the ones of LPC. We can see from figure 3 that the characteristic of a nasal is displayed in the changes of band and the disappearance of formants when the model is not really matched.

/ø/ is the most complicated nasal vowel. One reason for that is that the F'1 and F'2 are frequently very closed. We can not distinguish between them by increasing the order of LPC without any constraint, because there will be many unhelpful peaks, which we are not able to explain. Moreover the signal is discrete, therefore an excessive prediction will lead to ineffective calculation for harmonics. The zero-poles model just remedies this drawback. The evident difference prediction for formants by the two different models is taken as a good cue for recognizing the nasal /ø/; if we find a great difference for formants in the region lower 1400 Hz, then we take it as /ø/. It must be pointed out that it was only the results aimed at the phenomenon mentioned above. We can see that a simple decision on formants given by the two different models determines 50% of the /ø/, and the error ratio that an oral vowel is labelled as nasal is just 2%. (See the following table)

<table>
<thead>
<tr>
<th>vowel</th>
<th>5</th>
<th>a</th>
<th>e</th>
<th>e</th>
<th>a</th>
<th>o</th>
<th>u</th>
<th>y</th>
<th>ð</th>
<th>ø</th>
</tr>
</thead>
<tbody>
<tr>
<td>vowel labelled as /ø/</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

There is no great difference between /a/ and /ø/. The only difference is in the relative distance between F'1 and F'2. This distance is a little great for /a/ than for /ø/, in general this feature makes it possible for LPC to furnish an estimation for the frequencies of F'1 and F'2. In theory, there will be a Fin and A1n near F'1 and F'2, consequently the bands of F'1 F'2 and sometimes even F'3 obtained from LPC become too large. According to F'1 F'2, we can guess if a vowel is nasal. After that, we must observe changes of the spectrum for distinguishing between /a/ and /ø/. It will be /ø/ if the distance between F'1 and F'2 is reduced or has a tendency toward lower frequency range.
As for /œ/ and /œ/, it is easy to identify them just by the feature displayed in high frequency F1 and their important energy in frequency range 1400_1500 Hz.

In theory, all nasal vowels possess at least two nasal formants. One is near 500 Hz, and the other near 2000 Hz. We use the first nasal formant which deforms the first and/or second oral formants. As for the use of second nasal formant, there are still many problems. One of the important errors is how to differ the second nasal formant from the formant which is in reality caused by glottis.

CONCLUSION

We have observed in detailed the the French nasal vowels /œ/ and /œ/. Their nasal characteristics are evident in the frequency region below F'2. As for /œ/ and /œ/, we have not observed particular results, but we have found out that the nasal characteristics of these two nasals are displayed usually by high F'1. Furthermore, the energy between F'2 and F'3 is perturbed. The changes of F'1 and F'2 for /œ/ and /œ/ are evident, as while as F'2 and F'3 for /œ/ /œ/. Beside the analysis of the position of formants, we must pay attention to the mouvement of the spectrographic form.

REFERENCES