NASAL VOWEL TRANSFER FUNCTIONS : A STATISTICAL COMPARISON BETWEEN THE NASAL AREA AND THE AREA RATIO

Solange Rossato, Gang Feng
Institut de la Communication Parlée Université Stendhal / INPG
Domaine Universitaire 38040 Grenoble FRANCE
email : rossato@icp.inpg.fr

ABSTRACT

In most of studies on nasal vowel production, the nasal tract area at the uvula is used. The ratio between the nasal and the oral tract areas is also proposed. The aim of this paper is to determine which velar variable is more relevant of acoustic changes in nasal vowel transfer functions. It can be considered that each value of a velar variable corresponds to a probability distribution in the acoustic space. Samples of transfer functions representative of this probability distribution are generated by an articulatory model and the statistical Friedman-Rafsky test gives an objective criterion to the approximate notion of “separation” of two multivariate samples. Results show that when the nasal area increases, the sensibility of acoustic transfer functions to nasal area variations decreases and the area ratio becomes a more reliable parameter of acoustics changes in transfer functions.

1. INTRODUCTION

On the articulatory level, the main gesture that produces nasal vowels is the lowering of the velum, the oral tract becomes coupled with the nasal tract. This simple gesture connects the nasal fossa and the several sinuses to the oral tract. In constrast, acoustic transfer functions of nasal vowels are very complex, depending not only on the configuration of the oral tract, but also on the anatomy of the nasal fossa and on the position of the velum that controls the degree of coupling. Many studies tried to find the acoustic correlates of nasality, by modeling the production of nasal vowels. Nasality is supposed to increase with the nasal area at the extremity of the velum. Most of these agrees with a relative weakness of the first formant and proposes other secondary correlates. However detailed characteristics of the spectra of nasal vowels vary with the frequency of the first oral resonance and the magnitude of nasal coupling. If the oral vowels are well characterized by their formants, acoustic correlates of nasal vowels are not obvious. Furthermore, the nasal area is vowel dependent as the lowering of the velum is limited by the tongue position [1]. For example, with a high tongue position, the maximum nasal area should be relatively small because the velum hasn’t got lot of space when lowering. Feng proposed in [2] to use another parameter to avoid such vowel dependence, the area ratio *d*: the nasal area at the uvula divided by the sum of both the nasal and the oral tract areas at the uvula.

Acoustic to articulatory inversion has received considerable attention in recent years and for oral vowels, significant progress have been achieved concerning the estimation of the vocal tract shape from speech signals. Specific task for nasal vowels is first to recover a suitable parameter indicating the position of the velum or its evolution. In a first study [3], we took both variables in consideration and we recovered either the nasal area or the area ratio from synthesized transfer functions. Both variables were compared in terms of correct estimation rates.

In this paper, the purpose is to evaluate which velar variable is more representative of the changes produced in acoustic transfer functions of the vocal tract. The statistical multidimensional test called Friedman-Rafsky compares two data samples and accepts or rejects the null hypothesis that “ both samples are independent random samples from the same distribution”. Data samples are transfer functions obtained from an articulatory model which allows to control all parameters and overall the value of the nasal area or the area ratio. First, we present how data samples are generated for both variables, the nasal area and the area ratio. Then, the Friedman-Rafsky test is detailed and the sample size dependence is tested. Finally, after specifying the way the multidimensional test is applied, results are discussed and compared for both velar variables.

2. THE METHOD

In this section, we present the articulatory model and the theory of the Friedman-Rafsky test before using this statistical test to our specific problem.

2.1. Generation of data samples

The articulatory model. The use of an articulatory model allows the production of all the transfer functions needed and the control of the position of the velum. We consider that, when we fixe the value of the nasal area (or the area ratio) all the possible transfer functions cover a region of the acoustic space and their density is representative of the probability distribution corresponding to the given value of the velar variable. So, we must generate transfer functions with a same value of this velar variable.

The articulatory model is Maeda’s model based on X-ray images of a speaker (Patricia Barbier) pronouncing French sentences. Eight parameters regulate the jaw and tongue positions, the opening and the protrusion of the lips, the larynx length. A ninth parameter v_m is introduced to represent the velum position. All parameters are normalized, centered on zero which corresponds to the mean value of all the positions registered on X-ray images and with a standard deviation 1. For normal distributions, variations between –3 and 3 are supposed to cover most of the cases. The distribution of all the values of parameter v_m measured on X-ray images brings to the fore that this parameter does not follow a normal distribution. The lowest values are around –1 and they are obtained during plosive consonant production. When parameter v_m is between –1 and 0, the velar port is closed. The highest values are close to 4 and correspond to the rest position, nasal vowels or consonants. The default value 0 is used for the oral vowel production. So this articulatory model enables to get both oral area function and velar port area, areas being calculated from midsaggital sections. As data about Patricia Barbier's nasal fossa are not available, area functions of nasal tract and sinuses are given by Feng and Castelli in [2]. A set of nine parameters controls configurations of the vocal tract and corresponding area function. As this study only concerns vowels, constraints are imposed on the constriction area, the place of constriction and the lips area.

A velar variable: the nasal area. The nasal area variations when parameter v_m increases are studied for various configurations. A bijective relation exists between the parameter v_m and the nasal area except for high values of v_m (v_m > 3.5, that correspond to a nasal area value of 1 cm²) where there is a saturation which depends on the tongue position. The velum lowers until it rests on the tongue and stays in that position even if parameter v_m still increases. The nasal area values are calculated without taking into account such a saturation, because it depends not only on the parameter v_m but also on the eight other parameters. For example, for a nasal area of 2 cm², some transfer functions may have a nasal area value lower because of on tongue position, the oral area at the uvula should be null. Those transfer functions are still considered as having a nasal area value of 2 cm², as they are corresponding to the same value of parameter v_m.

Another velar variable: the area ratio. The other parameter proposed is the area ratio *d*, defined by

$$d = \frac{A_{nasal}}{A_{nasal} + A_{oral}}$$

When the area ratio *d* is 0, the velar port is closed. When the area ratio is 1, vocal tract is a single-tube tract from glottis to nostrils, the pharyngo-nasal tract. The velar port area is the area of the oral configuration at the extremity of the velum multiplied by the area ratio *d*. The oral tract area is multiplied by (1-*d*). Thus, the sum of both areas is constant.

Area functions can be calculated for a given value of the nasal area or of the area ratio. Then, 20 cepstral coefficients are computed from transfer functions. For a given value of the velar variable, samples of N vectors of 20 cepstral coefficients can be generated.

2.2. The Friedman-Rafsky test

This test was presented by their authors [4] as a multivariate generalization of the Wald-Wolfowitz two-samples runs test, whose null hypothesis is that the two univariate samples have the same distribution. In the univariate case, the statistic test is the number of runs of elements from the same population in the sorted list of pooled samples. In the multivariate extension, the minimal spanning tree (MST) [5] is used to "sort" the elements of the samples. Smith and Jain have used Friedman-Rafsky test to test uniformity against clustered alternatives [6] and multivariate normality of a data set [7]. An important propriety of this test is that there is no hypothesis made on the probability distribution : it can be uniform, normal, or unknown.

Let one sample of *N* data points be labeled X, and the *M* data points of the other sample be labeled Y. The MST of the pooled samples is computed using the Euclidean distance metric. The number of edges in the MST linking a data point from X to a data point from Y is called *T*. Considering that X and Y are independent random samples from the same distribution, the normalized parameter *T'*

$$T' = \frac{T - E[T]}{\sqrt{Var[T/C]}}$$

approaches the standard normal distribution when $N \to \infty$. The permutation parameter *C* is the number of edge pairs which share a common node and is calculated by

$$C = \frac{1}{2}\sum_{i=1}^{2N} d_i(d_i - 1)$$

where d_i is the degree of the i*th* node of the MST, and we have, with $L = M+N$:

$$E[T] = 2MN/L$$

$$Var[T/C] = \frac{2MN}{L(L-1)}\left[\frac{2MN - L}{L} + \frac{C - L + 2}{(L-2)(L-3)}\left[L(L-1) - 4MN + 2\right]\right]$$

The null hypothesis is rejected if $T' < Z(\alpha)$ where $Z(\alpha)$ is the α quantile of the standard normal distribution. For example, if $T' = -1.28$, then the null hypothesis is rejected with a significance level $\alpha = 0.10$. If $T' = 2.33$, then the null hypothesis is accepted and the significance level of that decision is 99.9 %. So, for two samples X and Y, we obtain a significance level of the null hypothesis.

2.3. Sample size dependence

To check the test for sample size dependence, we let N be 100, 500, 1000, 1500, and 2000. Data points are vectors of 20 cepstral coefficients. The sample X corresponds to the area ratio value of 0.5. The statistical parameter *T'* is calculated for samples X and samples Y of N data points when the area ratio

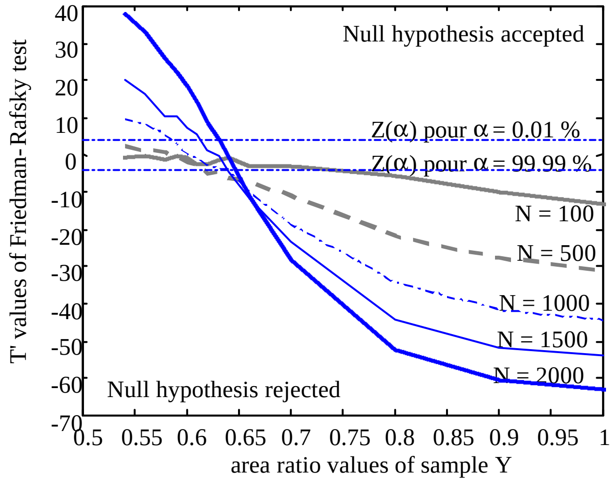

Figure 1: Statistic T' of the Friedman-Rafsky test between sample X (area value 0.5) and sample Y versus the area ratio of sample Y for different values of sample size.

increase. On Figure 1, we compare the variation of the statistical parameter *T'* versus the area ratio for different values of the sample size N.

When the sample size is limited, we have less information about the probability distribution, thus it is more difficult to accept or reject the hypothesis that both samples are issued the same distribution. For N = 100, knowledge on the probability distribution is not sufficient and the null hypothesis can not be accepted or rejected when the two samples are generated with the very different area ratio values. When N increases, the statistical parameter *T'* decrease monotonically with increasing the area ratio. The acceptation or the rejection of the null hypothesis becomes more significant. A sample size N=1000 seems to be a good compromise between the knowledge needed about the distribution and the calculation time.

3. RESULTS AND DISCUSSION

The Friedman-Rafsky test gives an objective measurement of the "separation" of two samples. The significance level is representative of the relative position of both samples. If they are mixed, the null hypothesis can not be rejected. If they are separated, the null hypothesis is rejected with a high significance level. So, the significance level can be used as a parameter which quantify the separation of both samples. In order to evaluate the changes in the acoustic space due to the velar parameter, we consider that changes are consequent when we obtain a given significance level α for two samples that only differ from their velar variable values. More precisely, for a given value of the velar variable ν, we determine the variation Δν as the statistic test between the sample X generated with the value ν and the sample Y generated with the value ν + Δν gives a significance level α. We noted that the significance level

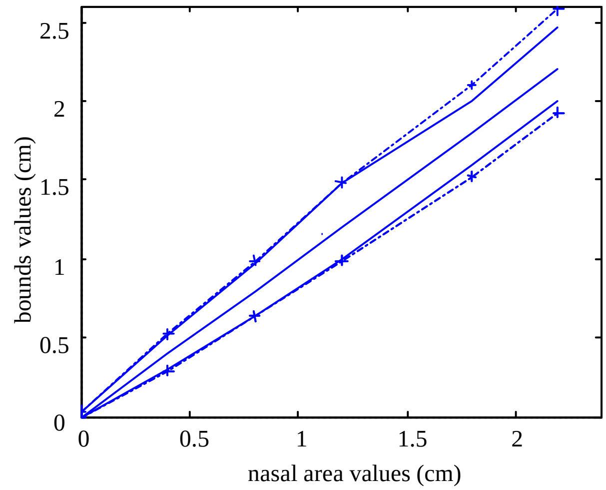

Figure 2: Upper and lower bounds versus nasal area for the two significance levels α=10 % (——) and α=1% (- - - -).

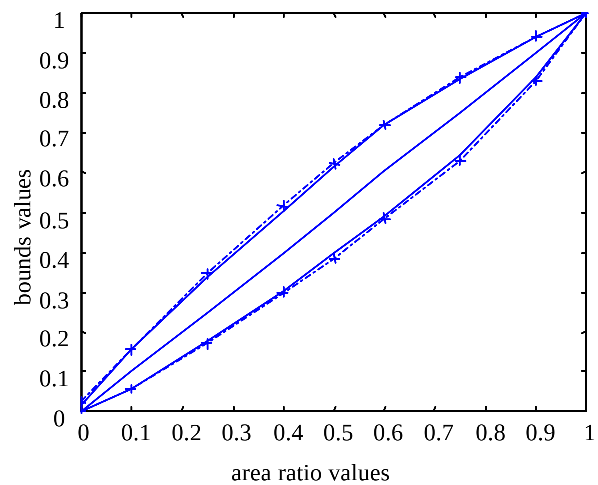

Figure 3 : Upper and lower bounds versus area ratio for the two significance levels α=0.10 (——) and α=0.01 (- - - -).

increases monotonically with the increment Δν when the sample size is sufficient. So, if the significance level is lower than the expected value, we decrease Δν, otherwise, we increase Δν. The value corresponding to the expected significance level is called the upper bound. The lower bound is obtained the same way by testing a sample X generated with the velar variable value ν and a sample Y corresponding to the value ν - Δν.

For various values of the nasal area, the upper and the lower bounds are plotted for α=0.10 and α=0.01 on Figure 2. The same way, Figure 3 represents the bounds calculated for various area ratio values and for the same significance levels. As shown for when studying sample size influence, bounds obtained for both significance levels are very similar either for the area ratio or for the nasal area.

For nasal area values close to 0, bounds are near the initial nasal area value, but when the nasal area increases, the upper and the lower bounds stay away from the nasal area value. A disadvantage of the nasal area is the tongue saturation, if the nasal area still increases, transfer function does not change because the velum can not go lower and rests on the tongue. So, when the nasal area is limited by the tongue position, transfer functions obtained are neglected during the statistical test. We only compares data points from articulatory configurations that have the same nasal area. Results can be interpreted in terms of acoustic changes due to nasal area against acoustic changes due to others articulators. For small nasal area, a little change in the nasal area produces important changes in transfer functions and both distributions are quickly considered as different ones. For higher values of the nasal area, a little change in the nasal area produces acoustic changes that are not sufficient to distinguish both samples as having different distributions. Acoustic changes due to an increase (or decrease) in the nasal area become less important comparing to changes due to variations in oral configurations for high values of the nasal area.

Considering the area ratio parameter, the upper and the lower bounds are close to the initial area ratio values. We can observe the same tendency as for the nasal area until the area ratio value reaches 0.5. Then, the reverse phenomenon is observed and bounds become more and more close to the initial area ratio when the latter increase. Acoustic changes in transfer functions due to the area ratio are more significant where the degree of coupling is small (oral area << nasal area or inversely) than when the degree of coupling is near its maximum (area ratio value of 0.5, oral area close to nasal area). When the degree of coupling is maximal, the dispersion due to the various oral configurations possible is maximal. Or, from another point of view, there are articulatory changes but acoustic transfer functions have a relative stability.

When the velum allows a small passage of air throw nasal cavities, both parameters seem to be correct and relevant of acoustic changes due to the position of the velum, they are both representative of the nasal constriction. When the velum goes lowering, the nasal area is less representative of acoustic changes due to the velum position than the area ratio: bounds are closer to initial values for the area ratio values than for the initial nasal area values. This qualitative comparison between both parameters confirms the results we obtained in a preliminary study: the estimation of small nasal areas was very good and that as the velum is lowered, the area ratio becomes a better parameter than nasal area. Acoustic spectra seem to be sensitive to velar port area constriction ; when the area increases, constriction disappears and the area ratio between the oral and nasal tracts is more representative of the changes in transfer functions. Quantitative comparison is not possible because both parameters are of different nature; the nasal area unity is cm^2 whereas the area ratio is without unity.

4. CONCLUSION

Relations between articulator movements and corresponding acoustics consequences are very complex. The statistical test proposed by Friedman and Rafsky allows to compare the influence of both velar parameters used in nasal vowel production studies. The area ratio and the nasal area seems to be equivalent and well representative of acoustics changes when the velum begins to lower. When the nasal area increases, the sensibility of acoustic transfer functions to nasal area variations decreases and the area ratio becomes a more reliable parameter of acoustics changes in transfer functions.

ACKNOWLEDGMENTS

The authors would like to thank Maeda for making available his model and Rafaël Laboissière for his insight and support in this study.

REFERENCES

[1] Maeda S. 1993. Acoustics of vowel nasalization and articulatory shift in French nasal vowels, in *Nasals, Nasalization and the velum* ou *Phonetics and phonology*, edited by M.K. Huffman et R.A. Krakow (Academic,New York), Vol 5, 147-167.

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

[3] Rossato S., Gang F. and Laboissière R. 1998. Recovering gesture from speech signal: a preliminary study for nasal vowels, in *5th ICSLP 98 Conference Proceedings*, Sydney, Australia.

[4] Friedman J. H. and Rafsky L. C. 1979. Multivariate generalization of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist*., vol 7, pp 697-717.

[5] Prim R. C. 1957. Shortest connection networks and some generalizations. *Bell Syst. Tech. J*. vol. 36, pp 1389-1401.

[6] Smith S. P. and Jain A. K. 1984. Testing for uniformity in multidimensional data. *IEEE Trans. Pattern Anal. Machine Intell*., vol PAMI-6, pp 73-81.

[7] Smith S. P. and Jain A. K. 1988. A test to determine the multivariate normality of a data set. *IEEE Trans. Pattern Anal. Machine Intell*., vol 10, pp 757-761.