Perceptual analysis of compensatory strategies in the production of the French rounded [u] perturbed by a lip-tube.

SAVARIAUX Christophe, PERRIER Pascal & SCHWARTZ Jean-Luc
Institut de la Communication Parlée, URA CNRS 368, INPG & Université Stendhal
46, avenue Félix Viallet, 38031 GRENOBLE Cedex 01, France
E-mail : {savariaux, perrier, schwartz}@icp.grenet.fr

ABSTRACT

In order to analyse the compensation strategies develop by 11 speakers to achieve, in spite of the lip perturbation, the perceptual goal we proposed two perceptual tests. Thus, the vowel [u] produced both in normal and perturbed conditions are perceptually evaluated. The results allow us to determine a [u]-region in the F1/(F2-F0) plane within which the produced sound is identified as a good quality [u].

INTRODUCTION

In a previous work we proposed an experiment the aim of which was to test the respective roles of the articulatory and acoustic levels in the control of vowel production. This experiment involved a lip perturbation impeding the usual articulatory strategy for the production of a rounded vowel [1]. To study how speakers are able to achieve their speech goal, in spite of the perturbation, a 20-mm diameter lip-tube was inserted between their lips while they produced the French rounded vowel [u]. Acoustic and X-ray articulatory data were recorded for eleven French native speakers for both normal and lip-tube conditions. Articulatory measurement in lip-tube conditions were made first immediately after the insertion of the tube (PF condition) and second after an "adaptation" procedure of 20 trials (PL condition).

The analysis of the results based on the F1/F2 comparison [1], showed that one out of eleven speakers (speaker OD) was able to acoustically compensate for the perturbation. For this aim, he moved his tongue backwards into the pharyngeal cavity. Remaining speakers showed a large variability in their compensation ability: four speakers (YP, ML, LJ, ML) presented no noticeable articulatory change and six others (MP, BC, CH, GA, JM, JY) displayed variable extents of tongue backward movements within the palatal region. Speaker OD's compensation abilities show that there is no physiological limitation to move the tongue backwards. Moreover, articulatory reactions, when they exist, were directed in the "right" direction, namely towards a compensation in the acoustic space. Thus speech production seems to be guided by auditory requirements.

However, for the majority of speakers the compensation is not completely achieved and the shape of the tongue remains close to what it is in the normal condition. One interpretation would be that speakers experienced difficulties in producing unusual articulatory configurations and hence resorted to the usual one. Another explanation could be found at the perceptual level. Indeed, the perceptual goal is probably not sufficiently characterised in the F1/F2 plane; in spite of an insufficient compensation in the F1/F2 plane, speakers could have reached their perceptual goal.

To refine the analysis of our data, we propose in this paper the results of two perceptual tests which have been achieved on both normal and perturbed utterances for all speakers.

THE IDENTIFICATION TEST

Method and procedure

The aim of the first test was to categorise the vowels produced in perturbed condition. These vowels were presented within a set of seven vowels delimiting the maximal vowel space of each speaker: the three extreme vowels [i, a, u] and four vowels acoustically located close to vowel [u], i.e. [o, œ, y, ø]. Each vowel was recorded in three conditions: one normal condition (N) and two perturbed conditions with a lip-tube (PL and P2). Thus, 21 stimuli are available for each speaker. All stimuli

were presented to 17 French adult listeners.

The procedure adopted for this test was as follows: in a sound-treated room, listeners listened to a stimulus by means of a headphone and then had to choose, without any time constraint, an answer on a monitor. The choice was made among seven possibilities [i, a, u, o, œ, y, ø]. Two seconds after the selection, another stimulus was sent to the headphone.

Results

Only results concerning the identification of [u] in N and PL conditions are presented here. For condition N and for all speakers, maximum identification (100%) was essentially obtained. The results for the PL condition are not so clear: seven speakers obtained a score near to maximum identification; 100% for 6 of them (CH, GA, LR, MP, OD, YP) and 94% (1 error among 17) for one (BC). For the remaining speakers, identification varies from 12% (speaker JM) to 0% for speaker JY (6% for speakers ML and LJ).

Interpretation

In condition N vowel [u] is clearly identified. We shall consider that an identification score smaller than 94% (more than one mistake) corresponds to a sensitive decrease of the vowel quality. Thus in the PL condition, 7 speakers have achieved, in spite of the perturbation, the required perceptual goal for vowel [u]. This somehow contradicts the analysis we proposed on the basis of acoustic data. We can observe an increase as high as 60% in F1 (speaker CH) or 46% in F2 (speaker LR) without change in vowel identification. However, when the F1 and F2 values become respectively higher than 400 Hz and 1100 Hz, the identification is no more correct. A large increase of F1 leads to a change of category from [u] to [o] (speaker LJ); simultaneous increases of F1 and F2 lead to confuse perturbed [u] with [œ] (speakers JY and ML).

The JM case clearly shows that F1 and F2 are not sufficient to understand in detail identification scores: the score was only 12% correct in spite of a formant pattern comparable to the one for speaker LR (F1 = 343 Hz, F2 = 851 Hz versus F1 = 344 Hz, F2 = 876 Hz) who obtained a maximum identification score.

The high perceptual scores obtained for 7 speakers could incite, in a first analysis, to minimise the conclusion made from the study of F1/F2 pattern: seven speakers and not only one, appear to be able to roughly compensate for the lip perturbation. However, a finer-grain analysis seems necessary if one looks at the speakers' strategy within the adaptation procedure. Indeed, the large formant pattern variability observed from trial to trial suggest that speakers, in spite of a correct identification score, were looking for a suitable articulatory configuration likely to produce a "better" quality [u].

THE EVALUATION QUALITY TEST

Method and procedure

To test such a hypothesis, we realised a second perceptual test the aim of which was to evaluate the vowel quality produced in the lip-tube condition for each speaker. For this objective, listeners had to rate the quality of the [u] on a scale from 1 to 7 (1 = very bad [u], 7 = very good [u]). The corpus was made of the three [u] of each speaker produced in conditions N, PL, P2 so a total of 33 stimuli have been used. Eighteen listeners participated to this second perceptual test, among whom sixteen had participated to the first one. The same procedure as in the first test was used. For this test, listeners randomly listened to each stimulus five times.

Results

First of all, the analysis made on the [u] produced in condition N shows that all the average rate are higher than 5 except for one speaker (JM; 4,09). Starting from this, we consider that a stimulus having an average equal or higher than 5 will be considered as a sound with the desired perceptual quality.

The average rating between listeners for all stimuli in the N and PL conditions are plotted in figure 1 in the F1/(F2-F0) plane. All frequencies are in Bark [2, 3], and a normalisation by F0 is adopted for F2 in reference to Traunmüller's studies [4].

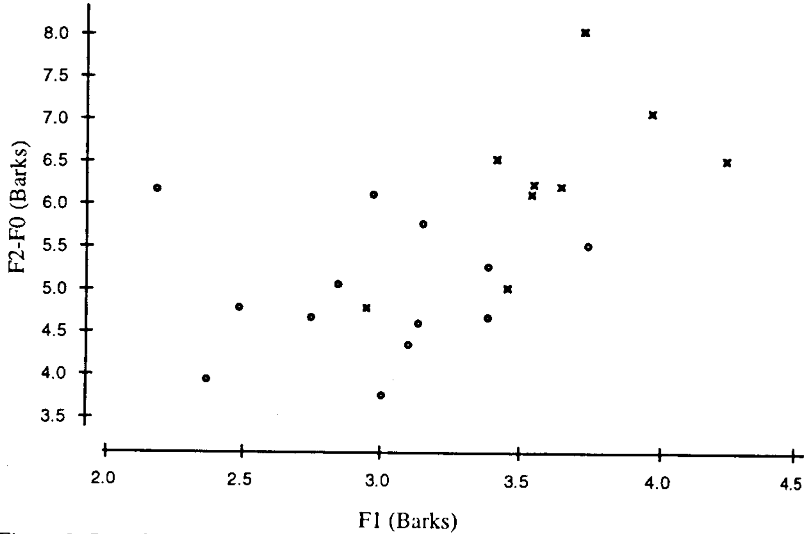

Figure 1: Distribution of the stimuli produced in the N and PL conditions for the eleven speakers. The stimuli having an average rate greater than 5 are represented by circle and those having an average rate smaller are represented by cross.

From this kind of representation, it is possible to propose a [u]-region within which a vowel has the required perceptual quality: almost all stimuli having an average rating greater than 5 are located in a region where F1 is smaller than 4 Barks and F2-F0 is smaller than 6 barks. The remaining well-rated stimuli correspond to F2-F0 slightly greater than 6 barks and F1 smaller than 3 barks. Note that stimuli produced in the PL condition by speakers CH, MP and GA are in this [u]-region. This tends to demonstrate that these 3 speakers have achieved the compensation in the perturbed condition. Note also that in this [u]-region 2 stimuli (JM in condition N and OD in condition PL) have an average rating less than 5. Of particular interest is speaker OD's production in the PL case: he produced a backward movement of the tongue leading to a complete compensation in the acoustic plane (attested by the position in the [u]-region); however, the perceptual effect is not sufficient (rate: 3.71).

Interpretation

These results shed light on the different strategies which are possible to compensate for the perturbation. Speaker CH did not produce a large reorganisation of the usual articulatory configuration for vowel [u]. Hence, his formant F2 remained fairly high, but this was compensated by a high F0 value leading to a correct F2-F0 value.

Speakers GA and in a lesser extent MP moved the tongue backwards inducing a decrease of F2-F0. This decrease was reinforced by a F0 increase though not sufficiently, but a low F1 value allowed to reach the desired perceptual quality.

As mentioned above, speakers JM and OD presented singular characteristics. In the case of speaker JM in the normal condition, the low [u] quality in spite of correct F1 and F2-F0 values show that one must consider all the spectrum to understand the perceptual effect of a sound. However, we shall concentrate further on the perturbed condition.

The case of speaker OD is more interesting. His F1/F2-F0 patterns in N and PL conditions are quasi-identical, but the perceptual evaluation falls down from 6.67 for condition N to 3.71 for condition PL. The F3 value provides the only significant difference between both conditions. This difference is then likely to play a decisive role in the perceptual evaluation. This hypothesis was verified with a simple perceptual test, in which the stimuli N and PL of this speaker were evaluated with and without a low-pass filtering at 1500 Hz. Results showed that in the PL condition the perception was better when the stimulus was low-pass filtered.

But if the perceptual product was insufficient, why did this speaker produce such a backward movement leading to a strong change of the tongue shape? The answer is given by the results of a last perceptual test, the aim of which was to understand speaker OD's behaviour through trials during the adaptation procedure. Results showed that at the beginning of the adaptation procedure, the [u] was confused with [œ] whereas at the end, the confusion was with [o]. Thus, the extent of the backward movement of the tongue during the adaptation procedure had a perceptually relevant effect: it allowed to reach, in the PL condition, a clear categorisation of the perturbed [u] as a velar vowel whereas before adaptation, confusion was possible between a central and a velar vowel.

CONCLUSION

First, this perceptual study confirms the importance of the perceptual goal in speech production. Speakers seem to have a clear representation of this goal and act in general in the right way to reach this goal in spite of the perturbation.

Moreover, this study proposes some interesting data useful for the understanding of this perceptual goal. It appears that in the F1/(F2-F0) plane, we can propose a [u]-region in the acoustic space within which the produced sound is identified as a "good" [u].

Finally, the study shows that in such a complex task where the perturbation imposes a complete reorganisation of the articulatory gestures, one can find: (1) subjects who do not need to compensate thanks to a correct initial configuration (speaker CH); (2) subjects who need to compensate and appear to know enough about the articulatory-perceptual relationships to be able to perceptually compensate up to a certain extent (speaker OD); and (3) subjects who should compensate and appear not to be able to do so (for example, speaker YP).

ACKNOWLEDGEMENT

The authors are sincerely grateful to all speakers and listeners for their contribution to these experiments.

This work is supported by the Esprit Basic Research Project n° 6975, Speech Maps.

REFERENCES

[1] Savariaux C., Perrier P. and Orliaguet J.-P. (in press), "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production", *Journal of the Acoustical Society of America.*

[2] Schroeder M. R., Atal B. S. and Hall J. L. (1979), "Objective measure of certain speech signal degradations based on masking properties of human auditory perception", in B.E.F. Linblom and S.E.G Ohman (Eds.), *Frontiers of Speech Communication Research*, (pp. 217-229). London: Academic Press.

[3] Zwicker E. and Feldtkeller R. (1981), *Psychoacoustique. L'oreille récepteur de l'information*, traduction française par C. Sorin, Paris: Masson et CNET-ENST.

[4] Traunmüller H. (1981), "Perceptual dimension of openness in vowels", *Journal of the Acoustical Society of America*, Vol. 69, pp. 1465-1475.