THE EFFECT OF MODIFYING FORMANT AMPLITUDES ON THE PERCEPTION OF FRENCH VOWELS GENERATED BY COPY SYNTHESIS

Anne Bonneau, Yves Laprie

LORIA-CNRS, Bâtiment LORIA, BP 239,
54506 Vandœuvre-lès-Nancy FRANCE.
Tel. (33) 3 83 59 20 80, FAX: (33) 3 83 41 30 79, E-mail:bonneau@loria.fr

ABSTRACT

With synthetic stimuli copied from natural vowels and including up to five formants, we investigated some transformations of the perceived identity of vowels by means of modifications of formant amplitude levels. Synthetic stimuli were generated by means of a new method of copy synthesis. With stimuli copied from /u/ we found that, despite the presence of F2, it is possible to transform the timbre of /u/ into that of a front vowel only by raising the amplitude level of F3 and higher formants. We also analysed how the timbre of the vowels /i/ and /y/ changed as a function of the level of F2, F3 and F4, and showed how sensitive was the timbre of the vowel /i/ to the decrease of the level of F3 and F4. Such transformations, realized with stimuli very close to natural vowels, reinforce the importance of formant amplitude levels in some vocalic distinctions.

1. INTRODUCTION

With two-formant synthetic stimuli, previous studies ([1, 2], among others) have shown that the variations of the relative formant amplitude levels affect the perceived identity of vowels. The most striking result of these experiments was the transformation of a high front vowel timbre into that of a high back vowel when the amplitude level of the second formant was drastically decreased.

With synthetic stimuli copied from natural vowels and including up to five formants, we investigated some transformations of the perceived identity of vowels by means of modifications of formant amplitude levels.

We used such quasi-natural stimuli in order to better predict what would happen if some formant region of natural vowels would be attenuated. This is of interest in the study of speech understanding by hearing-impaired people as well as in the study of speech in noise. From a theoretical point of view, we intended to confirm and even to reinforce the importance of formant amplitude levels in some vocalic distinctions.

We studied the transformation of the perceived identity of the back vowel /u/ into that of high front vowels, and that of the perceived identity of the vowels /i/ and /y/ into that of /y/ and /i/, respectively. For the stimuli generated from /u/, we tested the effect of the presence and of the absence of F2. This allowed us to better assess the role of F2 as well as the importance of the amplitude level of F3 in the perception of back vowels.

2. PROTOCOL

2.1 Corpus and stimuli

We chose one utterance of each of the three vowels /u,i,y/ from the French database BDSONS[3]. These vowels were uttered by a single French male speaker who had a low fundamental frequency (100-110 Hz) and a pronunciation which was both clear and well representative of the standard French. The duration of the selected utterances ranged from 160 ms to 200 ms.

The acoustic cues of these vowels served as references to generate the synthetic vowels /i,y,u/ which we wanted to be a close copy of the natural ones. All the other stimuli, with modified amplitude levels, were generated from these three synthetic vowels. We used a new method of copy synthesis which controls the parallel branch of the Klatt synthesiser[4]. The fundamental frequency as well as the amplitude and frequency of the first formants (three, four or five depending upon their visibility) were automatically evaluated by means of our copy synthesis algorithm (table 1). Our signal editor allows an interactive correction of the results by the user, but this step was not necessary. The acoustic cues listed above and other fixed parameters (formant bandwidths, overall gain…) constitute the input parameters of the Klatt synthesiser. By copying the evolution over time of the main vocalic information, we generated synthetic vowels whose quality is close to that of natural ones.

From the copy of each natural vowel, we created new stimuli only by increasing and/or decreasing the original level of one or more formants (formant frequencies were unchanged). The first formant was not altered. The modifications, which aimed at transforming the perceived identity of a vowel into that of another one, depended upon the couple of vowels considered. The limits of the amplitude

levels were set as follows: the upper limit corresponded to the amplitude level of the first formant; the lower limit was reached when the formant was no longer visible. In some cases, we have equally tested the effect of the absence of one formant. Some important modifications of the relative amplitude of two close formants could not be realized since they led to too prominent a spectral peak which "overwhelms" other formants in the spectrum.

We tried to transform the perceived identity of the back vowel /u/ into that of a front vowel, and the perceived identity of the vowels /i/ and /y/ into that of /y/ and /i/, respectively. For that purpose, we applied a series of modifications to vowel formant amplitude levels (table 2). All the possible combinations of the modifications were realized.

For the vowel /u/, we increased the level of higher formants. We did not decrease the amplitude level of the second formant which was not very prominent, but we tested the effect of its absence. Some stimuli generated from /u/ sounded ambiguous since they had the F1 of a close vowel and an important peak in the F2' area of the vowel /e/. We thus added two stimuli for which the very weak F4 and F5 were increased more than the F3 (by 20 dB) in order to favour the emergence of a non-ambiguous /i/ timbre.

For the vowel /i/, we decreased the amplitude of F3 and F4 and/or increased the amplitude level of F2. For the vowel /y/, we decreased the amplitude of F2 and F3, and/or increased that of higher formants. One stimulus, resulting from the (-20,+20) combination could not be created because of the spectral constraint mentioned before. Note that we always applied the same modification for F2 and F3 of the vowel /y/, and for F3 and F4 of the vowel /i/.

The corpus also contained nine stimuli generated from the vowel /ɔ/ but with modified formant amplitude levels. The analysis of the results for these highly ambiguous stimuli is beyond the scope of this study. Consequently, our corpus was made up of 36 stimuli (8 from /u/, 8 from /i/, 11 from /y/, 9 from /ɔ/).

2.2 The experimental task

Twenty native speakers of French (10 women and 10 men) served as listeners in this experiment. They all reported to have normal hearing. The subjects listened to the sounds in a quiet room.

They were asked to choose their response from among all the oral French vowels (/i,e,ɛ,a,y,ø,œ,u,o,ɔ/). Sounds which listeners were unable to distinguish were to be identified as unknown sounds. The subjects could give more than one response if they wanted to; the preferred answer -if any- was to be written first. A list containing the most frequent written form of each vowel followed by a monosyllabic word in which this form appears was given to the listeners before the test and commented. We insisted on the differences between mid-open and mid-close vowels sharing a same place of articulation, such as /e,ɛ/. These differences are not often clear in French speakers' minds although most can pronounce these sounds perfectly. The subjects wrote their responses on answer sheets designed for that purpose.

F	u		i		y	
F1	312	-	279	-	323	-
F2	732	-17	1895	-20	1783	-8
F3	2110	-34	3219	-15	1959	-9
F4	3190	-59	3553	-18	3208	-26

Table 1: Formant frequencies (Hz) and amplitude levels (dB) expressed in function of F1 level of the natural vowels which served as references to generate the synthetic stimuli.

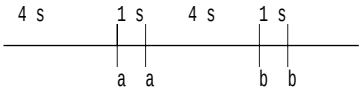
u		i		y	
F2	F3+	F2	F3+	F2-F3	F4+
id	id	id	id	id	id
abs	+10	+10	-10	-10	+10
	+20		-20	-20	+20
			abs		

Table 2: Modifications of the formant amplitude levels relative to the original levels (given in table 1). These modifications were applied to generate new synthetic stimuli. "Abs" stands for absence, FN+ means FN and higher formants.

The identification task included five test sessions, each of them with a randomised version of the test corpus. The responses were given after two consecutive presentations of the same stimulus. Each session was divided into blocks of twelve tokens with the insertion of a ten-second inter-block interval, a four-second interval between two different stimuli and an interval of one-second between stimulus repetitions. Figure 1 represents this temporal organisation. There was a two-minute break between each session. Ten tokens were presented before the first session in order to familiarise the listeners with the stimuli.

We thus collected five responses for each of the 36 stimuli and from each of the 20 speakers. Consequently there were 100 responses per stimulus.

The experiment lasted about 30 minutes.



Temporal organisation of a session. a and b represents two different stimuli.

Figure 1: Temporal organisation of the perceptual test

3. RESULTS

The results for each vowel are illustrated by figure 2.

3.1 Stimuli generated from the vowel /u/

The amplitude level of the F3 of /u/ (a peak located at 2000 Hz, which we will refer to as P2.0) appears to be an important cue to distinguish the high back vowels from the high front ones.

The relevance of this cue was brought to light by the responses to the stimuli which did not include the F2 of /u/ but included the F1, F3 (P2.0, the second peak for these stimuli) and higher formants of this vowel. The perceived identity of the stimuli changed from /u/, when the original levels of F3 (P2.0) and higher formants of /u/ were not modified, into /i,e/ when this level was increased by 10 dB.

These results are in good agreement with those of Ainsworth and Millar[1] as well as Chistovich and Chernova [2] which were obtained with two-formant synthetic vowels. When the first formant was set to about 300 Hz and the second one to 2000 or 2500 Hz, the subjects perceived the vowel /i/ for low L1/L2 ratios and the vowel /u/ for high L1/L2 ratios (L1 and L2 denote the amplitude level of the first and the second formant, respectively).

A stronger demonstration of the importance of the amplitude level of F3 (and higher formants) was given by the responses to stimuli which were very close to the natural vowel /u/ since they were made up of all its formants, F2 included. When the F3 amplitude level was not modified, the stimulus was perceived as /u/ in 98% of the cases, and /o/ for the remaining 2%. With an increase of at least 20 dB, the /i,e/ response became the most numerous. Consequently, it is possible to drastically modify the timbre of /u/ only by changing the level of F3 and higher formants.

A comparison between the results obtained for the two sets of stimuli generated from the vowel /u/, those which included its original F2 (first set) and those which did not (second set), showed that the presence of F2 favoured the persistence of the /u/ timbre when the F3 level was increased. The amplitude level of P2.0 had to be 10 dB higher in the first set than in the second set to change the perceived identity of the vowel. Then F2 which is not very visible and whose presence is not necessary to the perception of the /u/ timbre seemed to enhance the perceived identity of /u/.

For high F3 amplitude levels, an additional increase of the amplitude level of higher formants favoured the /i/-response to the detriment of the /e/-one, as expected.

To some extent, the knowledge of formant levels (measured by Pols et al. [5]) allows perception results to be predicted from amplitude modifications. Undoubtedly one of the best examples is that of /u,i/ because these two vowels share the same F1 frequency and the presence of one formant (F3 for /u/ and F2 for /i/) near 2000 Hz, F3 of /u/ being nearly 18 dB lower than F2 of /i/. Apart from the presence of F2, raising F3 of /u/ by 20 dB should approximately give the same formant level pattern as that of /i/. Indeed, this similarity did lead to the perceptual confusion between the two vowels.

3.2 Stimuli generated from the vowel /i,y/

The main energy concentrations of both vowels were located in similar frequency regions: a first formant around 300 Hz, a second energy concentration around 1900 Hz (P1.9) corresponding to the F2 of the vowel /i/ and the very close F2 and F3 of the vowel /y/, a third energy concentration around 3200 Hz (P3.2) corresponding to the F4 of the vowel /y/ and the very close F3 and F4 of the vowel /i/.

Previous experiments dealing with two-formant synthetic vowels showed that the best synthesis of the vowels /i/ and /y/ were obtained when the second peak (F2') was set to about 3000 and 2000 Hz, respectively [6].

It is thus very probable that a decrease of the amplitude of P3.2 would lead to the perception of /y/ for the stimuli generated from the vowel /i/ whereas a decrease of the amplitude of P1.9 would lead to the perception of the vowel /i/ for the stimuli generated from the vowel /y/. Our aim was both to confirm this hypothesis and to study the way it happened as a function of the amplitude changes.

Figure 2 shows that the perception of the vowel /i/ was particularly sensitive to the decrease of the amplitude level of P3.2 relative to that of P1.9. Thus a decrease of 10 dB altered the timbre of /i/; a further decrease (of at least 20 dB) led to the identification of the vowel /y/.

The timbre of /y/ appeared to be more persistent. Thus a 20 dB decrease of the amplitude level of P1.9 relative to the level of P3.2 was necessary to alter the perception of /y/ and to lead to some /i,e/ responses, while a drastic decrease (at least 30 dB) led to the perception of the vowel /i/. The high amplitude level of the F2-F3 energy concentration (-8 dB relative to that of F1) as well as the important bandwidth of this concentration could explain the persistence of the /y/ timbre.

Until now, we have analysed the results as a function of the modifications applied to the original amplitude levels of each vowel. Let us now analyse the same results as a function of the resulting amplitude levels of the second and the third energy concentrations (L2 and L3) of the stimuli -whatever vowel they were generated from. The stimuli were perceived as /y/ when L2 was higher than L3; they were perceived as /i/ when L3 was higher than L2 (from at least 5 dB, according to the few elements that were at our disposal) and the subjects hesitated between both vowels when L2 and L3 were close together.

4. CONCLUDING REMARKS

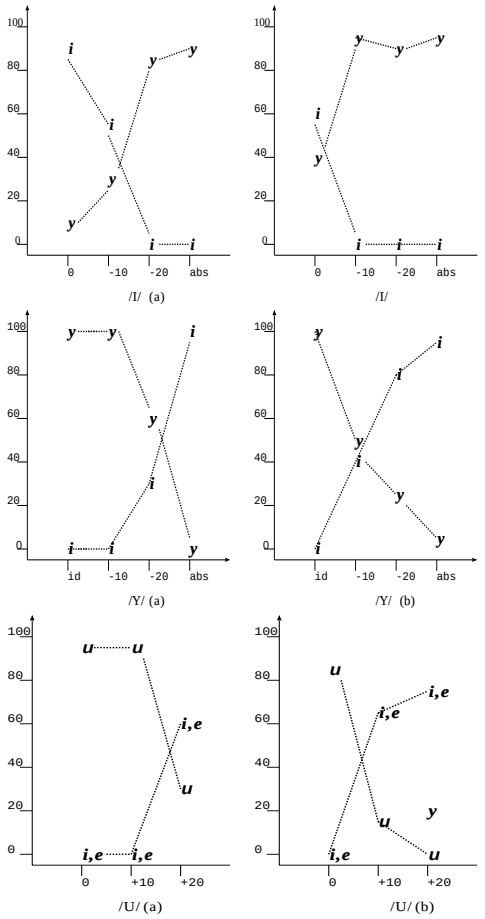


Figure 2: Percent of vowel responses (Y-axis) as a function of the modifications of the formant levels (X-axis). Negligible answers (less than 5%) are not included.
Vowel /i/: modifications of the amplitude level of P3.2 (F3-F4), (a): P1.9 (F2) is not modified, (b): P1.9 is increased by 10 dB.
Vowel /y/: modifications of the amplitude level of P1.9 (F2-F3), (a): P3.2 (F4) is not modified, (b): P3.2 is increased by 10 dB.
Vowel /u/: modifications of the amplitude level of P2.0 (F3 of /u/), (a): the original F2 is present, (b): the original F2 is absent

We presented some transformations of the perceived identity of quasi-natural vowels by means of modifications of formant amplitude levels. With synthetic stimuli copied from the vowel /u/ and including all the formants of this vowel, we showed that it is possible to transform the timbre of /u/ into that of a front vowel only by raising the amplitude level of F3 and higher formants. We also analysed how the timbre of the vowels /i/ and /y/ changed as a function of the level of F2, F3 and F4. and showed how sensitive was the timbre of the vowel /i/ to the decrease of the level of its third and fourth formants.

Such transformations, which were realized with stimuli very close to natural vowels reinforce the importance of formant amplitude levels in some vocalic distinctions.

5. REFERENCES

[1] W.A. Ainsworth and J.B. Millar. The effect of relative formant amplitude on the perceived identity of synthetic vowels. *Language and Speech*, 15:328–341, 1972.

[2] I.A. Chistovich and E.I. Chernova. Identification of one- and two-formant steady-state vowels: a model and experiments. *Speech Communication*, 5:3–16, 1986.

[3] R. Carré, R. Descout, J. Mariani, M. Eskenazi, and M. Rossi. The French language database: defining, planning, and recording a large database. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1984.

[4] Y. Laprie and A. Bonneau. Une méthode de synthèse par copie pour piloter le synthétiseur de Klatt. In *Journées d'Etude sur la Parole*, Martigny, Switzerland, Juin 1998.

[5] L.C. Pols, H.R. Tromp, and R. Plomp. Frequency analysis of dutch vowels from 50 male speakers. *J. Acoust. Soc. Am.*, 53(4):1093–1101, 1973.

[6] R. Carlson, G. Fant, and B. Grandström. Spectral peak detection and centres of gravity effect. In G. Fant and M.A. Tatham, editors, *Auditory analysis and perception of speech*, pages 55–82. Academic Press, New York and London, 1975.