Contributions of F1 and F2 (F2\') to the perception of plosive consonants

René Carré\textsuperscript{1}, Pierre Divenyi\textsuperscript{2}, Willy Serniclaes\textsuperscript{3}, Emmanuel Ferragne\textsuperscript{4}, Egidio Marsico\textsuperscript{4}, Viet-Son Nguyen\textsuperscript{4}

\textsuperscript{1} Laboratoire Dynamique du Langage, CNRS-Université Lyon 2, Lyon
\textsuperscript{2} Speech and Hearing Research, Veterans Affairs Northern California Health Care System, Martinez CA, USA
\textsuperscript{3} CNRS-LEAPLE, Université René Descartes, Paris
\textsuperscript{4} Centre MICA, CNRS/UMI2954, Hanoi University of Sciences and Technology

Abstract

This study examined the contribution of F1 and F2 alone on the perception of plosive consonants in a CV context. Applying a 3-Bark spectral integration the F2 frequency was corrected for effects of proximity either to F1 or to F3, i.e., was replaced by F2\'. Subjects used a two-dimensional Method of Adjustment to select the F1 and F2 consonant onset frequencies that led to a subjectively optimal percept of a predefined target CV. Results indicate that place prototypes are guided by F2 and are largely independent of F1. Nevertheless, while F2 alone is sufficient for segregating place for some consonants and vocalic contexts, it is insufficient for explaining the perception of place.

Index Terms: speech perception, plosives, F2', 3-Bark integration

1. Introduction

Various perception experiments have shown that the third formant is essential for differentiating the place of articulation of consonants like \(/d/\) and \(/g/\) and that the slopes of the first and second formant are insufficient to convey this information. Harris et al. [1] have noted that although, in synthesized \(/CV/\) experiments, the consonant \(/d/\) cannot be obtained in the vocalic context \(/i/\) without the third formant, it can be perceived in the \(/ae/\) context. Similarly, Godfrey et al. [2] reported that the rate and/or the direction of third formant transition was necessary for perceiving the distinction between \(/d/\) and \(/g/\), whereas for the \(/bl/-/dl/\) contrast the transition of the second formant was sufficient. In duplex perception, Mann et al. [3] also used F3 for the \(/ld/-/gd/\) distinction. In contrast, Delattre et al. [4] were able to produce synthetic CV perceived as \(/bw/\), \(/dv/\), or \(/gv/\) by changing the slopes of the first two formants alone. However, for \(/gl/\), \(/g\ell/\) and \(/g\alpha/\), they observed a different acoustic behavior compared to \(/gl/\), \(/g\ell/\), \(/g\alpha/\) and \(/g\alpha/\) [5].

In production, Öhman [6] represented the \([Cu]\) transitions in the F2/F3 plane with different preceding vowels and identified three main regions, each corresponding to one of the three consonants \(/b/\), \(/d/\), \(/g/\). Lindblom [7] displayed the F2 and F3 onsets of the plosives as a function of the second formant of the final vowel taken from Öhman’s data and observed that, indeed, there is a boundary between \(/b/\) and \(/d/\), \(/g/\) in the F2-onset/F2-vowel plane and another boundary between \(/g/\) and \(/b/\), \(/d/\) in the F3-onset/F2-vowel plane. The different behavior observed by Delattre et al. for the perception of \(/g/\) was also observable in production as two different locus equations for back and front vowels [8].

However, following another deductive approach [9], the closed-open Distinctive Region Model (DRM, see Figure 1) can predict the vocal tract regions corresponding to place of articulation of \(/b/\) (region R8), \(/d/\) (region R6) and \(/g/\) (region R5). F3 is necessary to make the \(/dl/-/gd/\) distinction based on a simple, increasing or decreasing F-pattern shape. While the VCV tokens examined by Öhman [6] were produced using the model via superimposition of a consonantal gesture on vocalic gestures [10], the role of the first three formants is defined in a more direct fashion by the DRM. In the case of \(/Cu/\) (and \(/Col/\), \(/C\alpha/\)), however, the DRM configuration becomes closed-closed symmetric, with the consequence that the regions deduced will be different from those of the closed-open variety illustrated in Figure 1.

But what are the consequences of these two configurations to the production of CV syllables? Are F3 transitions an absolute necessity allowing perceptual distinction of CVs in the F1-F2 plane alone? The objective of the research reported in this paper was an attempt to answer this question.

Figure 1: Variations of the first three formants from a closed-open DRM model corresponding to closing-opening of a specific region.

Relying on Lindblom’s [7] findings, the major problem is expected to arise in the case of \(/g/\) in all vocalic contexts, especially that of \(/au/\). We should recall that at the onset \(/g/\) is produced with F2 and F3 at consonant onset is about 200 Hz, as illustrated by the formant trajectories in Figure 2. The slope of the F2 transition is small at the \(/g/\) onset due to the proximity of F2 and F3 (about 200 Hz) but it becomes steep near the vowel \(/au/\). While in the traditional spectrogram representation the F2 onset
maximum is around 1800 Hz and leads to a small F2 transition slope, taking into account the 3-Bark integration [12] or the F2′ effect [13] (Figure 3), F2 increases to about 2000 Hz, which leads to a steeper F2 transition slope. Such a slope could signal the /d/g distinction even in absence of F3. The experiments described in the following sections tested the notion that allowing the formants to be broader by applying the 3-Bark integration to the process, as suggested by [13], it is possible to generate CV tokens with a formant synthesizer driving only two formants and for listeners to perceive the tokens as intended. We were especially curious to see what F1-F2 trajectories led to a percept of /gV/ and /Cu/. To test our hypothesis that CV tokens generated by the first two formants lead to veridical percepts, we chose a two-dimensional version of the Method of Adjustments, i.e., a paradigm in which the subject selected the F1 and F2 onset frequencies of a given CV target.

2. Perception experiments

Synthesized CV tokens were generated in real time with V being the vowel [a], [i], [u], or [neutral] and C the stop consonant [b], [d], or [g]. Formant frequencies F1 and F2 for /a/ were 750 and 1200 Hz, for /i/ 250 and 2500 Hz; for /u/ 250 and 750 Hz, and for the neutral vowel 500 and 1500 Hz. The murmur was 20 ms, the CV transition 80 ms, and the Vowel 150 ms in duration. There was no burst at the beginning of the transitions. In each trial, a target CV was specified on the subject’s screen that also displayed a graphical representation of the F1-F2 plane, as illustrated in Figure 4. By placing a cursor controlled by the computer mouse on this F1-F2 graph, the subjects could control the F1 and F2 frequencies at the onset of the consonant. The synthesizer generated a CV token with a linear transition between the selected consonant F1 and F2 values and those of the target vowel. The subjects’ task consisted in changing the consonant onset frequencies until the resulting CV token sounded “best”, i.e. until it led to a subjectively optimal percept of the specified target. There were ten trials for each combination of the four consonants and three vowels. The order of the 120 trials was random. Fifteen Francophone subjects with phonetic experience took part in the experiment conducted in a quiet room using headphones.

2. Results

The mean F1 and F2 frequencies of the /b,d,g/ prototypes are presented in Figures 5 and 6, respectively. Examination of Figure 5 suggests that differences between place prototypes are not related to F1. This is confirmed by a Place × Vowel repeated measures ANOVA run on Bark transforms of the prototypical F1 frequencies. The Vowel effect was significant (F(3,39)=8.4, p<.001) but neither the Place effect nor the Place × Vowel interaction were significant (both F<1). Examination of Figure 6 suggests that differences between place prototypes are indeed related to F2. This is confirmed by a Place × Vowel repeated measures ANOVA run on Bark transforms of the prototypical F2 frequencies. Both the Place effect, the Vowel effect and the Place × Vowel interaction were significant (both F<1). Examination of Figure 6 suggests that differences between place prototypes are indeed related to F2. This is confirmed by a Place × Vowel repeated measures ANOVA run on Bark transforms of the prototypical F2 frequencies. Both the Place effect, the Vowel effect and the Place × Vowel interaction were significant (both F<1). Discriminant analyses run on the 12 pairwise place contrasts, separately in each vocalic context, with the F2 prototype frequency as predictor of place categories gives Percent Correct Classification scores ranging from near chance level (52 %, for bu/gu) to 100 % (for ba/da). Obviously, F2 is not universally sufficient for identifying place categories.

1 For /d/ F2 is higher than generally observed due to the proximity of F3. As a result of the 3-Bark integration, F2 thus becomes F2′.
Additional analyses examined the dynamics of the formant transitions. From the results of each subject on each trial, we calculated the length of the trajectory (cf. Figure 4, the length of the red line) and its direction (the angle of this line with the F1 axis). The data were subjected to an ANOVA with either the trajectory length or the angle as the dependent variable. Interestingly, all analyses showed highly significant subject, vowel context, and consonant main effects as well as vowel-consonant interaction (all p<0.0001). The data were also analyzed using a generalized linear model (GLM) with trajectory length and direction as joint criterion variables and subjects, vowel, consonant and the product of vowel and consonant as the predictors. Univariate F tests for the subject effect was more significant for trajectory direction than for length (F(1,1551)=3.87 [p<0.05] and 31.36 [p<0.001]) and the vowel effect was highly significant for both the length and the direction (F(1,1551)>10 [p<0.001]). However, the consonant effect was not significant for trajectory length (F<1 [p>0.8]) but highly significant for trajectory direction (F(1,1551)=24.79 [p<0.0001]). Similarly, the Vowel × Consonant interaction was not significant for trajectory length (F<1) but significant for trajectory direction (F(1,1551)=3.92 [p<0.05]). These results suggest that the subjects were not likely to base their choice of F1 and F2 transition onset frequencies on the extent of the trajectory but were highly likely to choose formant frequencies lying on the slope of their preferred transition. In other words, they were likely to adopt transition velocity as the cue on which they based their decision.

Looking at the subject-by-subject F1-F2 starting frequency averages of blocks of 10 trials for the three consonants in the four vocalic contexts it becomes clear that the 15 subjects employed different strategies and that these individual differences may have been masked by the above reported analyses of the complete data set. To illustrate the different strategies subject appear to have adopted, Figure 7 shows results of two subjects for the three stop consonants in the /a/ context, and Figure 8 shows results of two subjects for consonants in the /u/ context. While, on the surface, in all vocalic contexts the subjects’ performances appear to differ mainly with regard to the dispersion of the trial average blocks, the /u/ context unveils deeper individual differences. In fact, just by looking at the clustering of data points for each of the three consonants (shown in different colors) one notices that the order of the clusters along the F2 axis is not the same for all subjects. Such ordering discrepancy might be explained by different subjects listening to different spectral cues when the consonant precedes the /u/ vowel – a vowel having F1 and F2 energy concentrated below 1 kHz. Alternatively, the differences between individuals might be due to differences in the decoding of the F2 transition in the absence of F3.

4. Conclusions

The present results show that place prototypes are related to F2 but not to F1. However, in a dynamic view, F1 might still
Contribute to the percept by defining a transition slope. Although F2 alone is sufficient for separating place prototypes for some contrasts and in some vocalic contexts it is not entirely sufficient for explaining place perception. Obviously, other acoustic cues are needed and previous experiments show that F3 plays an important role (for recent evidence see [14]).

5. Acknowledgements

We would like to thank the ANR (contract ANR-2010-BLAN-1916-03). This work is also supported by Veterans Administration Research.

6. References


