PRODUCTION AND PERCEPTION OF VOWEL CATEGORIES

René Carré ¹, Willy Serniclaes ² & Egidio Marsico ³

¹ ENST-CNRS, Paris, France
² LEAPLE CNRS, Villejuif, France
³ DDL-CNRS, Lyon, France
rene.carre@enst.fr

ABSTRACT

It is well known that isolated vowel perception is non-categorical: “if we conceive of a continuum between the ideals of categorical perception and continuous perception, we find that stop consonants are closer to the categorical ideal and vowels are closer to the continuum ideal” (Repp, et al., 1979). But, such a statement was based on perceptual tests with stable vowels and mainly for vowel continua located on the [ai] trajectory. In a preceding study (Carré, 2004), the area function space and the corresponding acoustic space were already shown to be dynamically structured following the predictions of the DRM model of speech production. Privileged formant trajectories were as follows: [ai], [ay], [au] for the [a, e, i] [a, œ, ø, y] and [œ, o, u] vowel sets respectively. The present study examines further evidence in support of the DRM model, based on data pertaining to the production and perception French vowels. Results will be discussed in terms of vowel representation as sub-products of vowel-to-vowel trajectories.

INTRODUCTION

In preceding paper (Carré, 2004), formant trajectories were automatically deduced from acoustic theory applied on a 18cm length tube. Two main criteria were used: maximum acoustic contrast and maximum formant variation for minimum area function deformation. This deductive approach yields places of articulation and phonological systems both corresponding quite well to those observed in speech production. The results obtained by the deductive approach can be modeled by the Distinctive Region Model (DRM) (Mrayati, et al., 1988). This model structures the acoustic and area function spaces; it displays privileged vowel trajectories in the F1-F2 plane, on which vowels can be represented. This means that, according to the deductive approach, vowel trajectories appear first, and then isolated vowels. This result is worth to study because it can be interpreted at the production and perception level: vowel trajectories could be observed directly or indirectly in production and represented as such in perception. Vowels could not be independent static exemplars but lie on these trajectories. In this paper, different measurements and perceptual experiments are proposed to study these hypotheses for French. The nasal vowels are not taken into account.

VOWEL TRAJECTORIES

The vowel trajectories intrinsic to an acoustic tube structured by the DRM model are represented in Figure 1. They structure in a simple way the acoustic space and are the results of the deformation gestures of the model. It should be noted that these gestures are closely correlated to the muscular effects of the genioglossus and styloglossus. The figure represents the phonological capabilities of the acoustic tube when this tube is efficiently deformed, i.e. when minimum area deformation involves maximum acoustic variations. The phonological
capacities are in terms of trajectories, not in terms of static privileged locations. Vowel categories lie on vowel trajectories in the F1-F2 plane and are then the consequences of vocalic trajectories. The phonological trajectories are obtained by one or two phonological gestural deformations of the model: one constriction gesture (with consequences in terms of cavities) and one lip gesture. In French, 3 trajectories are selected among the set of possible trajectories: [ai] obtained with only one deformation gesture controlling the main constriction from back to front by transversal control; [au] (or [çu]) also with only one deformation gesture controlling both the place of the constriction and the degree of lip opening; and [ay] which can be described as [ai] gesture plus labial gesture or as only one deformation gesture controlling both the place of the constriction and the degree of lip opening.

Figure 1. Vocalic trajectories obtained with the DRM model with the corresponding vowels. Dotted lines are labialised trajectories.

The set of vowels obtained with these three trajectories are [a, e, i, e, o]. [α, ɔ, o, u], and [a, ə, o, y].

VOWEL PRODUCTION
Two different sets of results are examined here in relation with trajectory components: one on formant characteristics pronounced in isolation and the other on vowel sequences in spontaneous speech.

Isolated vowel characteristics
The 11 French oral vowels were pronounced, randomly, with computer control according to the following example: “As in the sentence: Voilà le lapin, say: lapin, la, a”. Four French subjects (3 males and 1 female--sm) took part in the experiment. The first two formants of the isolated vowels were measured during all the production of these quasi stable vowels. The formant variations of each of the vowels are represented separately for each subject in Figure 2. The formant variations are fairly similar and privileged directions are observed pointing generally to
[a] as hypothesized. Individual differences can be observed: For (rc) the trajectories are more stable than for the other three speakers and the trajectory [y, ø, œ]. points to [a] instead of [a]. (em & sm) do not produce separated [a, ø]. (sm) does not produce separated [ø, œ].

![Isolated Vowels (rc)](image1)

![Isolated Vowels (em)](image2)

![Isolated Vowels (ws)](image3)

![Isolated Vowels (sm)](image4)

Figure 2. Representation in the F1/F2 plane of the vowel formant frequencies for four subjects. The formant variations correspond to the evolution from the beginning to the end of only one production.

<table>
<thead>
<tr>
<th>Vowel succession</th>
<th>Number of occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ae]: 8515</td>
<td>[ea]: 5525</td>
</tr>
<tr>
<td>[eœ]: 5068</td>
<td>[ii]: 4978</td>
</tr>
<tr>
<td>[eœ]: 5068</td>
<td>[ee]: 4686</td>
</tr>
<tr>
<td>[ia]: 4667</td>
<td>[ae]: 4612</td>
</tr>
<tr>
<td>[aa]: 4584</td>
<td>[ie]: 4459</td>
</tr>
<tr>
<td>[æœ]: 3361</td>
<td>[ai]: 4283</td>
</tr>
<tr>
<td>[æ]: 3119</td>
<td>[ ea]: 3018</td>
</tr>
<tr>
<td>[iœ]: 3011</td>
<td>[âœ]: 2907</td>
</tr>
<tr>
<td>[âœ]: 2892</td>
<td></td>
</tr>
</tbody>
</table>

Vowel successions
The structuring in terms of vocalic gestures can also be observed from speech production: assuming syllabic coproduction (Kozhevnikov and Chistovich, 1965; Öhman, 1966; Fowler, 1992), V1 and V2 should be somehow related in V1CV2 utterances. Table 1 shows that the
most frequent V1V2 successions in spontaneous French production lie on the [ai] trajectory (Carré, et al., 1995).

PERCEPTUAL EXPERIMENTS

Experiment 1. Categorical vowel perception
Categorical perception on both [ie] and [iy] continua were compared. In the F1-F2 plane, 8 equal intervals in Bark scale were defined: [ie] (3 intervals) and [iy] (5 intervals). The corresponding synthetic signals were used for identification (25 presentations for each sound) and discrimination (25 presentations – distance maximum between two items: + or – one interval). The four preceding subjects took part in the experiment. Figure 3 shows the results: for (em & sm) the observed discrimination scores are close to those predicted from the labeling data, thereby indicating that are almost fairly categorical; the two other subjects (rc & ws) exhibit different degrees of over-discrimination. Notice that the amount of over-discrimination tends to be larger between /i/ and /y/ than between /i/ and /e/. Another perceptual difference between the two continua, is that the /i-e/ phonemic boundary is fairly stable across individuals (as indicated by the location of the predicted discrimination peaks, around S9 in Fig.3), whereas the /y-i/ boundary varies between S4-S5 and S6-S7 (Fig.3).

![Graphs showing categorical perception for different subjects](image)

Figure 3. Predicted and observed categorical perception between isolated /y, i, e/.

Experiment 2. Perceptual vowel representation
The four preceding subjects took part in a perceptual experiment in which they had to find out with the use of a formant synthesizer the best vowel exemplars of the 11 French oral vowels by
adjusting F1 and F2 in a graphical F1-F2 plane representation (Johnson, et al., 1993). These vowels were produced with a preceding neutral vowel (F1=500, F2=1500, F3=2500Hz). The durations of the neutral vowel, the transition and the vowel to adjust were respectively 100, 100 150 ms. Twenty five sets of values were obtained and are represented in Figure 4 (in Bark scale). Most of the vowels stand on the two main trajectories [ai] and [au] but it is not the case for (sm). The labialised vowels [y, ø, œ] do not however lie on the [ay] trajectory. This result can be explained by the fact that two phonological gestures contribute to the [ay] gesture: the [ai]-constriction gesture and the labial gesture, thereby generating two dimensions of variability.

Figure 4. Representation in the F1-F2 plane of the best vowel exemplars for 4 subjects.
DISCUSSION

Our results show overall that there are definite trends for vowel categories, both in production and perception, to follow the trajectories deduced from acoustic theory and for vowel boundaries in perception to be more stable when located along these trajectories. The production data (Fig.2) show that, for the four subjects, the isolated vowels vary along directions pointing to either [a] or [α]. At the perception level, Experiment 1 shows that perception is more categorical and the phonemic boundary is more stable for a continuum following a gestural trajectory, namely [i-e] located along the [i-a] trajectory, than for one located aside, namely [i-y] continuum which is located outside the [i-u] trajectory. Experiment 2 results indicate that, for three among the four subjects under study, the vowel category vary along the predicted [ai] and [au] trajectories. The results of the 4th subject (sm) are much less clearcut, which might be explained by the fact that she is bilingual speaker (Arabic French). Notice however, that this deviant behavior is only found for perception. Another difference between production and perception is that, for all subjects taken together, the labial vowels follow the [ay] trajectory at the production level but are more dispersed at the perception level. Finally, still another discrepancy between the productive and perceptual representation of vowel categories lies in the size of the vowel triangle, which is larger in the perceptual representation (compare Figs. 4 and 2), as already observed by Johnson et al. (1993). To summarize, these preliminary results show that, at production and perception level, the [ai] and [au] trajectories have a specific role which is has now to be further characterized. It is necessary to increase the number of subjects to bring out the subject characteristics from the main behavior.

ACKNOWLEDGEMENTS

This work is partly financed by a grant from "Action concertée Incitative: Complexité" of the French Ministry of Research

REFERENCES


