Production and perception of Vietnamese vowels

Eric Castelli*, René Carré**

*International Research Center
Hanoi University of Technology, 1 Dai Cô Viêt , Hanoi, Vietnam
Eric.Castellisimca.edu.vn
**ENST-CNRS, Rue Barrault, Paris, France
rene.carre@enst.fr

Abstract

It is well known that vowels can be produced in isolation, acoustically stable in such a way that they are represented as points in the F1-F2 plane. But, in a preceding study [1], following the predictions of the DRM model of speech production, the area function space and the corresponding acoustic space were already shown to be dynamically structured. Privileged formant trajectories were as follows: [ai], [ay], [au] and [aui] for the [a, e, i] [α, ã, ãο, y] and [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 of Vietnamese vowels. Results will be discussed in terms of vowel representation as sub-products of vowel-to-vowel trajectories. Two specific Vietnamese vowels ‘ả’ & ‘â’ will be studied.

1. Introduction

In a previous paper [1], 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 deformation of the area function. 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) [2].

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 isolated vowels are then on the trajectories. This result is worth studying 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 Vietnamese vowels. It is often affirmed by Vietnamese linguistic researchers that there exist only 9 Vietnamese vowels. It is often affirmed by Vietnamese linguists that there exist only 9 Vietnamese vowels. It is often affirmed by Vietnamese linguists that there exist only 9 Vietnamese vowels.

2. 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 trajectories are closely correlated to the muscular effects of the genioglossus and styloglossus. The Figure 1 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 Vietnamese, 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 [aui]) also with only one deformation gesture controlling both the place of the
constriction and the degree of lip opening; and [au] which can be described as [u] gesture without labial gesture or as only one deformation gesture controlling the place of the constriction with fixed marge lip opening. The set of vowels obtained with these three trajectories are [a, e, i, j], [a, ã, o, u], and [a, y, u]. But what about Vietnamese vowels ‘â’ & ‘á’? Are they specific vowels which in this case we could name temporally [á], [ý]?

3. 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. The 11 Vietnamese oral vowels were pronounced 5 times, randomly, with computer control according to the following example: “As in the word: la, say: la, a”. The Vietnamese words were: ‘lăn’, ‘lã’, ‘larga’, ‘lê’, ‘li’, ‘lo’, ‘lọ’ ‘lọ’, ‘lu’ & ‘lư’, which correspond respectively to [lân], [la], [ý], [lr], [le], [lí], [lo], [ýl], [ło], [lu], [łu]. The vowel ‘á’ (íá) does not exist in open syllable. Five Vietnamese subjects (5 males: Anh, Son, Quang, Hung and Hong) took part in the experiment.

![Figure 2: Representation in the F1/F2 plane of the vowel formant frequencies for two male subjects. The formant variations correspond to the evolution from the beginning to the end of only one isolated vowel production.](image)

3.1. Isolated vowel characteristics

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 for two subjects and one production in Figure 2. The privileged directions obtained by deduction with the DRM model are observed pointing generally to [a] as hypothesized. These directions were also observed in CV production by [3] called inherent spectral change. But two sets of vowels are more or less acoustically closed in the plane F1/F2: [í/a], and [ý/ý]

On the acoustic point of view and in a first approximation we could say that these vowels are respectively identical. However, Vietnamese speakers observed that it is difficult for them to pronounce [í] and [ý] in isolation (static mode with a long duration): it seems that these vowels are always pronounced in a dynamic mode. It is interesting to observe that, since all other vowels are pronounced in isolation with a monotonous tone (flat pitch evolution), these two vowels are always pronounced with a non monotonous tone (rising tone 2): it seems that, because speakers are forced to pronounce these vowels in a static mode, they continue to add a dynamic aspect with this changing tone (see Figure 3).

![Figure 3: Tone (pitch) of the vowels [í] and [a] pronounced in static mode (subject Anh).](image)

3.2. Vowel characteristics in [l] context

The three sets of vowels, corresponding to our three main trajectories are shown in Figure 4a, b and c. Again privileged directions are observed at the end of the vowel production. But in the plane F1/F2, it is also difficult to classify [í/a], and [ý/ý/ú]. However, it is interesting to note that the [ý] trajectory finishes while returning horizontally (F2 constant) whereas there is no final consonant as in [lán] and this point consists in a common property of these both vowels. We remember that vowel ‘á’ does not exist in open syllable.

In [4], we proposed to take into account the slope (rate) of the CV transitions to discriminate vowels lying on the same trajectory. Figure 5 a) and b) show the corresponding results. The slopes S are calculated from smoothed formant variations (from three successive formant values, Fsi=(Fi+Fi+1+Fi+2)/3; the distance between two successive value is 12ms, and Sİ=Fsi+1− Fsi). Indeed, the vowels can be discriminated according to the maximum slopes of the positive peaks. For the 5 productions of subject Anh, the means (standard deviations) of the maximum slopes of the CV transitions for F1 (in Hz/ms) are the following:

- [í/a]: 122Hz/ms (18) / 97Hz/ms (9)
4. Vowel perception

We wished to test the situation of these two particular vowels at the perceptive level. In a first part, the experiment consisted in searching the synthetic vowel corresponding best to the vowel of the previous word, by clicking (pointing) in the F1/F2 plane. Each vowel is required 5 times. We think that these selected vowels then correspond to their internal representation [5]. Then, we will reproduce the experiment with the same vowels preceded by the consonant [b].

4.1. Perception of isolated vowels

Figure 5 shows the results for 4 subjects. Standard deviations might make us think that the vowels have a great number of possible choices on the F1 axis, thus on our basic trajectories. But surprisingly, a distinction between vowels [a/a] and [y/y/uu] is possible starting from the static formantic values of the sustained vowels. The small standard deviations for [a/a] and [y/y/uu] make us think that a static representation exists for these two vowels although the dynamic aspects observed in production can be represented by these points in F1/F2 space.
4.2. Perception of vowel in context

There is no important difference between isolated vowels and vowels in context [b] (we can compare subject Hong in figures 6 and 7). We found again small standard deviation for [â] and [ tô], which means that these vowels are clearly distinct. There is some move of the positions of the two vowels but with no significance (other classical vowels are moving too).

5. Discussion & perspectives

Production of Vietnamese vowels shows the privileged directions obtained by deduction with the DRM model (in isolated mode and in [l] context mode). In production Vietnamese vowels ‘ă’ and ‘â’ present the same acoustic characteristics (formants F1 and F2) that ‘a’ and ‘o’ ([a] and [ô]). However, it seems in a first approximation that ‘ă’ and ‘â’ present dynamic characteristics: non monotonous pitch and, in the [l] context mode, the vowels both present trajectory finishes while returning horizontally. If we take into account the slope (rate) of the CV transitions, the two vowels are clearly distinct because their maximum slopes are bigger than for the other vowels. In perception, we found that the 11 Vietnamese vowels (‘ă’ and ‘â’ too) are clearly distinct. However, while ‘ă’ seems to be positioned on =#?, the position of ‘â’ is less precise depending of subjects. Moreover, in order to precise with more details this “dynamic hypothesis” of these specific Vietnamese vowels, we plan to continue our experiment by studying these vowels in different consonant contexts and also with different tones (the production of the syllable with complex tones - like the broken tone - could modify its dynamic characteristics).

6. References

[1] Carré, R. "From acoustic tube to speech production," Speech Communication 42, 227-244, 2004