ARTICULATORY-ACOUSTIC RELATIONSHIPS IN UNVOICED STOPS: A SIMULATION STUDY

SHINJI MAEDA

Dept. RCP
Centre National d´Etudes des Telecommunications
22301 Lannion, France

ABSTRACT

An articulatory model for consonant-vowel (CV) syllables, where C=[p, t, or k] and V=[i, a, or u], was formulated in terms of vocal-tract (VT) area function. Listener identification functions indicated that C with a high score (100 %) can be synthesized by manipulating two articulatory parameters, the "position" along the VT length and the "shape" of the occlusion. The acousic effects of these two parameters are manifested from the burst onset to the vowel transition. The consonant identity can be predicted reasonably well on the basis of the presence or absence of two spectral attributes for the burst, in a context-independent manner. Why the burst alone can predict the consonantal place? The reason is that the effects of, particularily, the shape can be manifested concomitantly on the attribute of the burst and on formant (F-) transitions of the vowel, both signaling a specific consonant. It is suggested then that the listner´s processing indeed exploits cues distributed on the sound stream from the burst to vowel transitions, in which the context-independent attributes for burst may serve as an "anchor" in the identification.

INTRODUCTION

In a previous paper [1], we have described the mechanism of the VT excitation during the unvoiced stop release. The source sound, of cause, undergoes a particular spectral modification which is interpreted as a specific consonant by listeners. In the past decades, a great deal of research was accumulated in search of acoustic cues that specify place of articulation for stops. Two types of methods were employed; acoustic analysis of natural tokens to find out the acoutic correlates and examination of listener´s responses to synthetic stimuli in which acoustic characteristics are systematically manipulated.

This paper describes yet another approach which consists of, first, the formulation of an articulatory model for the CV-tokens. Second, informal and formal identification tests followed to determine essential model parameters for producing the CV-syllables with a high quality. Finally, the acoustic manifestations of such parameters are examined closely by means of spectral analysis or of calculation of the VT transfer functions. If a particular manifestation can explain the listener identification of consonants, it can be considered as a good candidate for the cue actually operating in the listener´s processing.

AN ARTICULATORY MODEL FOR THE CV-SYLLABLES

We assume a heavily anticipated articulation of the vowel during the preceding consonant. The VT area function defined by a piecewise-constant function is fixed to its configuration for the vowel, except in the vicinity of the occlusion, where the cross-sectional areas expand with time after release. We consider the following three different types of closure "shapes": i) Labial (L)-type; the area expansion is limited to a single section corresponding to the supraglottal closure as shown in Fig. 1a. We assume that this shape represents the release gesture by the participation of the lips. ii) Dental (D)-type; the closure section and the five sections directly behind it expand in such a way that a relatively smooth connection

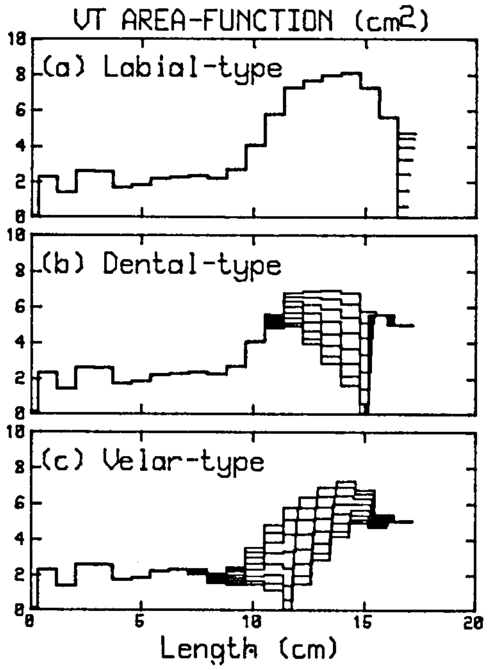

Fig. 1 Time-varying area functions for the three different closure types, sampled at every 20 ms following the release. For all cases, the target vowel is [a].

between the closure section and cavity behind (back-cavity) is maintained through release, as shown at Fig. 1b. D-type represents the dental gesture by the tongue apex and blade. iii) Velar (V)-type; The directly connected sections in both back and front cavity expands with the closure section, as presented at Fig. 1c. V-type is intended for the velar (or palatal) gesture involving the dorsum.

The way of the expansion of the closure section is specified by an exponential rise function. The cross-areas of the sections in the front and back of the occlusion also expand exponentially, but its onset rise is smooth without discontinuity. The smooth rise was necessary to prevent multiple excitation at vicinity of the occlusion. Resultant time-varying area function is fed to an acoustic VT simulator [2] for synthesis or for VT transfer calculations.

IDENTIFICATION TEST

In preliminary experiments, the three stops [p, t, and k] were synthesized with a reasonable quality by appropriately varying the value of the two articulatory parameters (the "position" along VT length and the "shape", L-, D-, or V-type), while the other parameters, such as VOT (=25 ms), glottal and release dynamics, were kept to fixed values. CV-tokens, therefore, were prepared by varying systematically the position, n, from 1 to 9, where the closure is located at the n-th VT section from the lips (n=1), and by varying the shape, for the three different target vowels, [i, a, and u]. The stimuli were randomized with five repetitions of each token for each closure shape.

Three experienced and three naive listeners participated in the test. Each listener was asked to identify the consonant, [either p, t, or k], and then type in the corresponding key on a computer key-boad. The next token was presented to the listener 1 sec after the response. The listeners were provided also a repeat request option upon which the same token is repeatedly presented. The total number of the repeats for each token was used in the interpretation of data as a measure indicating a quality of consonants.

The identification score for [p, t, and k], as a function of the position, for the target vowel [a] is presented for L-type at Fig. 2a, for D-type at Fig. 2b, and for V-type at Fig. 2c. The number of repeat requests for each token is also plotted in Fig. 2 by the dashed lines. As expected, the position of the occlusion is an essential factor in the production of the stops. Notice, however, that [t] cannot be produced by L-type (at Fig. 2a). The score for this consonant is only 37 % at best. For the same position (n=4), with D-type (at Fig. 2b) or with V-type (at Fig. 2c), the score reaches 100 %. For the velar [k], the score with L-type (at Fig. 2a) is relatively high, about 80 %. The number of repeat requests, however, is great, about 10 times, indicating an uncertain quality of the sounds as [k]. On the contrary, with V-type (at Fig. 2c), the identification function for [k], in fact for all three consonants, exhibits an ideal "categorical" response. Notice that the number of repeats, shown by the dashed line, increases only at the phonetic boundaries or at the extreme position (i.e., n=9).

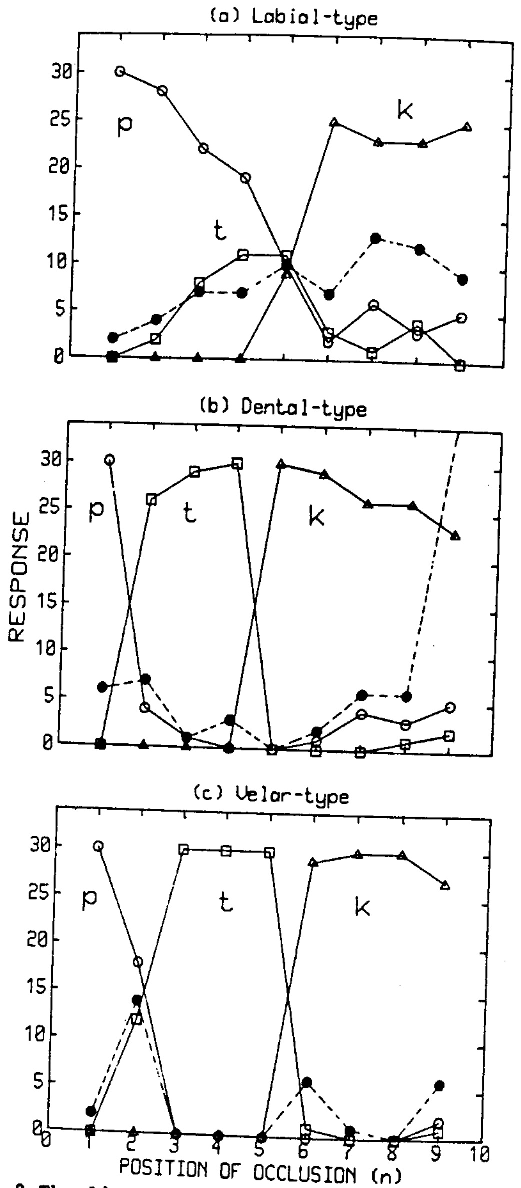

Fig. 2 The listener's responses as a function of the positions for the three different closure types. The full score, i.e. 100 %, corresponds to 30 on the ordinates (six listeners times five repetitions). The dashed lines indicate the total number of repeat requests by the six listeners for each token.

For the other two target vowels, [i and u], the responses were similar to the case of [a] described just above, except labial-dental contrast in [i]-context. The consonant [p] with a high score (90 %) was produced only with L-type, while [t] with score 100% was only with D-type, for both at the position n=1.

It became clear that not only the position, but also the shape of the occlusion is an essential factor for the consonants to be identified correctly. The acoustic manifestations of these two articulatory parameters, therefore, must be relevant to the listener's processing.

ARTICULATORY-ACOUSTIC RELATIONSHIPS IN BURST

The concomitant acoustic effects of the position and the shape of the occlusion are manifested in various form through out from the burst onset to the vowel transition. We shall attempt to sort out the acoustic characteristics that are consistent with the listener's reponses.

Due to the fact that during release, the aperture of the closure section is relativly small and the VT excitation sources are located at the exit of the closure, the acoustic characteristics related to only the front cavity appear on the burst [3]. The acoustics of the front-cavity is specified by the length (and thus by the position) and by how the cross-sectional area varies along its length. The listener's responses indicated that the dominant parameter for the velar, [k], is the position. Then, the spectral attribute of burst signaling the velar must be related to the front cavity, more specifically, the resonances of that cavity.

From the identification test, it is the shape that is more critical for the labial-dental distinction. If the aperture of the occlusion is, say, greater than 0.2 cm^2, the coupling effects appear on the transfer function and thus on the spectrum, regardless of the shape. When the aperture after release is still less than , say, 0.2 cm^2, the difference in the shape can manifest in the degree of the acoustic coupling between the front and back cavity. For a CV-syllable with L-type, due to the strong area-function discontinuity at the inlet of the occlusion (see Fig. 1a), the coupling is minimal. For the same CV, but with D- (or V-) type (see Fig. 1b), a smooth connection of the back-cavity to the occlusion is maintained from the onset through release. The coupling, therefore, is considerably enhanced in comparison with L-type.

In our synthesis, the rate of the area expansion at release was fixed to 20 cm^2/s. The presence or absence of the coupling effect, signaling the shape (L-type or D-type), therefore, can appear on the spectrum of the burst within 10 ms following the release. In other words, if the burst spectrum contains a series of peak and dip pairs (corresponding to that of pole-zero pairs in the transfer function) which is the indication of the coupling, then the shape is D-type or V-type. The absence of such spectral attribute implies the L-type.

The burst spectrum, therefore, can contain rich information to determine the position and the shape, which is essentially vowel-context independent. It is noted however that the effects of the two parameters can appear on the vowel transition, which are context dependent. In the following sections we shall examined a qualitative correspondence between the consonant identity (i.e., place) and spectral attributes of burst, and vowel transition for particular cases.

SPECTRAL ATTRIBUTES OF BURST AND PLACE

Attribute Pole-Zero for dental [t]

Burst spectra of the synthetic CV's, where the shape is D-type, and the target [i], are shown in Fig. 3. The position is varied from n=1 (closure at the lips) presented at the top in Fig. 3, to n=9 (at the posterior extreme) presented at the bottom. For the calculations, an 8 ms half Hamming window was aligned to each release onset.

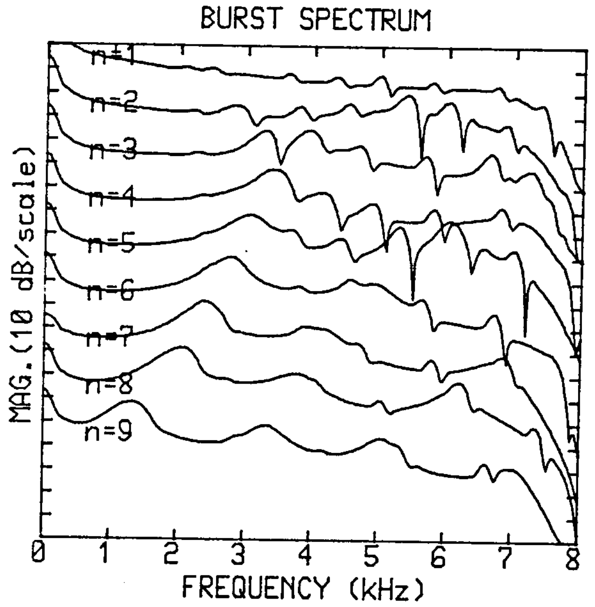

Fig. 3 **Burst onset spectra for nine different positions, from the lips (n=1) to the posterior extreme (n=9). The burst signals were synthesized with the target vowel [i] and the D-type closure.**

For the position n=1 shown in Fig. 3, the front cavity is absent, resulting in a "falling" spectrum, which might imply, therefore, the labial [p]. Indeed, Blumstein and Stevens [4] has proposed the "diffuse-falling" gross shape of the burst spectrum as an invariant property for labials. The identification function indicated, however, that this token scored 100 % as dental [t]. Notice that the spectrum (n=1) exhibits a series of peak-dip pairs, which is a typical signature of the presence of the coupling, and thus of D-type shape. Let us call this kind of spectral characteristics attribute "PZ". The identical CV token, except L-type instead of D-type, scored 90 % as [p]. In this case, the attribute-PZ was absent. It is then stated that the shape is more critical than the position in labial-dental distinction, at least, in this particular case.

When the position is located at a slightly posterior position in the VT, i.e., n=2 or 3, a broad peak, BP, (at around 5.5 kHz for n=2 in Fig. 3) appears as the resultant effect of the front-cavity resonance. Thus BP might be considered as the attribute signaling dental, since the position is appropriate for the dental [t]. The presence of BP at a high frequency results in a rising mid-frequency spectrum, which may correspond to the invariant property "diffuse-rising" for dentals [4]. In our data, however, attribute PZ predicted more consistently the listener's identification of the dentals with a high score. An explanation for this will be described latter.

Attribute Prominent-Peak for velar [k]

At the position n=4 or greater, the length of the front cavity becomes relatively long, and the resonance frequencies shift toward lower frequencies and can exhibit a prominent peak, PP, as seen in Fig. 3. The presence of PP was common to the burst spectrum identified as [k] with a high score. The invariant property "compact" for velars [4] may correspond to the attribute PP. The presence of PP means a long front-cavity, and then an appropriate position as velar.

It should be mentioned that the CV tokens corresponding to n=7, 8, and 9 were considered as , at best, an ambiguous stop by the listeners and not velar, even though the skewed but prominent peak is present at low frequencies in each burst spectrum. The skewed peak is due to the rapid shift of the corresponding free-pole toward higher frequencies as the aperture of the constriction expands following release. When the shape, V-type, is employed for the otherwise identical token, the prominent peak is shifted toward a high frequency by more than 1 kHz, and indicated a less skewed peak. This is, of cause, due to the effects of the narrowed front-cavity toward the occlusion. Except for the extreme position (n=9) the corresponding tokens scored 100% as [k]. An inspection of the spectrograms shows that the shape, V-type, places the burst prominence at a "right" frequency in relative to the F-pattern in the vowel transition. In the case of the tokens with a high score, or in particular, of natural tokens, an appropriate position for velar implies the appropriate shape, and then the prominence at the right frequency. Consequentry, the presence of the attribute PP alone would suffice to specify velar.

Rules for predicting place

The prediction of place for unvoiced stops from the burst spectrum became evident. First, the presence of the attribute PP signals velar [k]. If absent, it means the position is for dental or labial. Therefore, if the attribute PZ is present, the consonant is dental [t], since it must be produced with the shape, D-type. If absent, then it is the labial [p].

THE SHAPE AND FORMANT TRANSITIONS

We shall concentrate our attention to the question why the shape is critical in the contrast labial vs. dental. A pertinent example was found in tokens with the target vowel [i]. As mentioned before, for the same position, n=1, token with L-type is identified as [p], wheares that with D-type as [t]. The spectrograms of the two tokens are shown in Fig. 4. Observe that F3 (the third formant) for the L-type at Fig. 4a is clearly rising, while F3 for the D-type at Fig. 4b is slightly but lowering. It has been demonstrated for voiced stops [5] that the F3 transition plays an important role in their identification, that is consistent with the effect of the shape described here. A similar effect of the shape on F2 transition was observed for the target vowel [a], (i.e., [pa vs. ta]).

It can be stated, then, that the influences of the shapes, L- or D-type, are manifested coherently on the burst and on the F-transitions, assuring a robust identification. This is the reason, probably, why the attribute PZ is favored over BP.

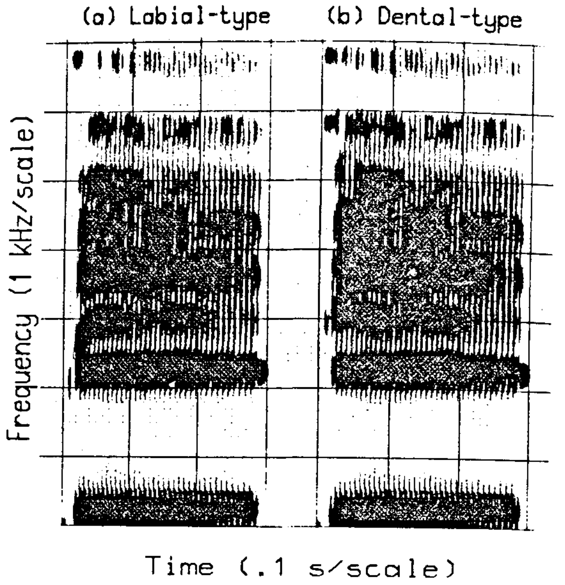

Fig. 4 Spectrograms of two CV syllables where C is identified as [p] at (a), and as [t] at (b). For both cases, the closure position is at the lips (n=1), and the target vowel is [i].

CONCLUDING REMARKS

In informal listening, it was often found that the identification of the stops upon signals corresponding to the burst alone or to the vowel part alone was difficult or impossible. When they had been assembled forming normal CV-tokens, however, the consonant was easily identified. From such experience, it is tempted to speculate that a suppression and/or an enhancening mechanism over the distributed cues are operating in the listener's processing. Coherent attributes found in both burst and vowel transition enhance each other. On the contrary, inconsistent attributes are suppressed. Such mechanisms may explain the listener identification functions in more comprehensive way.

REFERENCES

[1] S. Maeda, "Une source d'excitation coherente dans les occlusives," 15e JEP GALF, Paris, 43-46, 1985.

[2] S. Maeda, "A digital simulation method of the vocal-tract system," Speech Communication, 1, 199-229, 1982.

[3] G. Kuhn, "Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance," J.Acous.Soc. Am., 65(3), 774-788, 1979.

[4] S. Blumstein and K. Stevens, "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J.Acous.Soc.Am., 66(4), 1001-1017, 1979.

[5] K. Harris, H. Hoffman, A. Liberman, P. Delattre and F. Cooper, "Effect of third-formant transitions on the perception of the voiced stop consonants," J.Acous.Soc.Am., 30(2), 122-126, 1958.