

# RELEVANT SPECTRAL INFORMATION FOR THE IDENTIFICATION OF VOWEL FEATURES FROM BURSTS

*A. Bonneau*

CRIN-CNRS & INRIA Lorraine, 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

This paper presents a first attempt to extract relevant spectral information for the identification of the following vocalic context from French stop bursts. For this purpose, we studied the acoustic spectra of bursts used in a perceptual experiment which showed that listeners were able to identify vocalic features from bursts [1]. The corpus was made up of stimuli of 20-25 ms duration extracted from natural monosyllabic words which combined the initial stops /p,t,k/ with the vowels /i,a,u/. The low frequency limit of the frication noise as well as the frequency of the most prominent peak of the burst appeared to be very interesting cues for the identification of the vocalic context. Using these cues, most of contexts (/i/ from /t,k/, /u/ from /p,k/ and /a/ from /p,t/) have been very well classified without specification of the consonant.

## 1. INTRODUCTION

Because of extensive stop-vowel coarticulation [4] the acoustic spectra of stop bursts depend a great deal upon the identity of the following vowel. Using natural stimuli made up of either the onset of stop-vowel syllables or the entire burst of voiceless stops, previous experiments [7,3,6,1] showed that listeners were able to recover the main features of the following vowel from bursts. More precisely, the results of all these experiments concurred to show that the subjects could very clearly identify the front-back feature and the high front characteristics of vowels /i/. We also found that the proximity between the places of articulation of the stop and of the following vowel led to a clear identification of the vocalic context. The aim of this study is to investigate the relevant spectral information for the identification of vocalic features from (French) stop bursts. For this purpose, we studied the acoustic spectra of the bursts whose vocalic context was well identified by listeners. Given our small database, we were only concerned here by the most relevant cues.

## 2. PERCEPTUAL EXPERIMENT

In a previous experiment [1], we investigated the perception of the vocalic context from stop bursts (see section 3 for a description of the stimuli). The experiment

comprised two conditions: a training condition (subjects listened to training stimuli and simultaneously read their identity) and a lack of training condition (not related in [1]). The identification rates obtained under training condition were as follows: /pi/: 63%; /ti/: 90%; /ki/: 86%; /pa/: 85%; /ta/: 94%; /ka/: 77%; /pu/: 91%; /tu/: 37%; /ku/: 99%. The lack of training affected these three contexts significantly: /pu/: 67%; /ta/: 79%; /ka/: 33%.

## 3. METHODS

We used stimuli from previous experiments devoted to the perception of French stop bursts [1,2]. The corpus was made up of 90 stimuli extracted from natural CVC and CV monosyllabic words which combined the three initial voiceless stops /p,t,k/ with the three vowels /i,a,u/. Each syllable was uttered twice by five French male speakers. The duration of the stimuli was 25 ms. In order to cut off all traces of vocalic segment, bursts whose duration was too short (one dental and twelve labial bursts) were lengthened. For this purpose, we duplicated the burst tail end. The preparation of the stimuli was done with the help of the Snorri signal editor [5].

For our acoustical study, we segmented the burst into two parts: the transient and the fricative segment (the aspirative segment is generally absent from French stop bursts). For this purpose, we used an automatic procedure conceived by Y. Laprie [2]. The boundary between the two events corresponds to the instant when the global similarity of the large band burst spectrogram with the two average spectra is maximized. There sometimes was a slight gap between the “ideal” limit and the proposed one and we corrected the results manually. The two parts of the burst can be seen in Fig.1 which displays an example of a burst spectrogram for each CV context.

## 4. ACOUSTIC CUES

### 4.1. Cue for the /i/-context

In some contexts (/ti,ki,tu/), the center of gravity of the burst spectrum rose from the transient to the fricative noise and reached very high values (table I). Fisher-Jorgensen (1954) [4] noted this phenomenon for the Danish /t/ which is affricated: “the noise following the explosion of /t/ has exactly the same frequency as a Danish

s”. It appeared that the low limit of the friction noise was particularly high in the context /i/, especially for the most intelligible contexts (/ti,ki/). We roughly defined this limit in the following manner: *frequency from which the energy becomes higher than the mean energy of the burst spectrum*. In order to eliminate a peak in low frequencies observed sometimes for /pi/, the region below 700 Hz was not taken into account.

It seems that the low frequency limit of the friction noise depends upon the starting point of formant transitions (table I). In general, the determinant formant was F2, but it could also be F3 for /ti,ki/ contexts and F1 for /a/ context. This criterion appeared to be very relevant for the identification of /i/ from /t/ and /k/ (see Fig. 2 and Fig. 3). All /ti,ki/ items but one had values above 2000 Hz, while all non-/i/ contexts but one (a /a/ from /k/, identified as /i/<sup>1</sup> under lack of training condition) had values lower than 1800 Hz. This criterion was less relevant for the (less intelligible) context /pi/. Actually, values for well perceived /pi/ were only slightly higher than values for /tu/ and overlapped values for five /ka/, the vocalic contexts of which were often identified as /i/ under lack of training condition. We observed that values for labials followed by /i/ were clearly higher than values for labials in /a,u/ contexts.

We found a positive correlation ( $r = 0.8$ ) between the values of the cue and the number of /i/ responses for each item. If we eliminated the /pi/ stimuli, the correlation jumped to 0.84.

#### 4.2. Cue for the /u/-context

Listeners clearly identified /u/ from labials and velars. From dentals, only three /u/ contexts were identified as /u/, and the best rate was 62%. The difference between the energy at the F2 frequency region of /u/ (we choose 700 and 1300 Hz as limits for the region) and the energy in high frequencies separated /pu,ku/ contexts from all contexts but /pa/, confounded with /pu/ because of the falling spectrum of labials, particularly visible in /a,u/ contexts. The spectrum of labials followed by /a/ had a prominent peak in very low frequencies, not observed for /pu/ contexts, which was due to the anticipation of the following open vowel (see Fig. 2).

We thus defined the following criterion, calculated on the fricative segment of the burst:

$$2 E(700,1300) - (E(0,400) + E(1600,2500)),$$

where  $E(x,y)$  is the highest energy value observed in the frequency region delimited by  $x$  and  $y$  expressed in Hertz. It was possible to increase the value of the last frequency limit (2500 Hz) until at least 3000 Hz without modifying the behavior of the cue.

This criterion clearly separated all the /pu,ku/ contexts, which had the highest values, from other contexts. Only one item (a labial in /u/-context) had a value clearly outside the characteristic /u/ region, while one /pa/ stood at

its border (see Fig.4).

To correctly classify the three /u/ which were well identified from dentals, it would be necessary to raise some limits in the preceding formula; this modification would be responsible for few errors of classification for /pu,ku/ contexts. A non-invariant criterion, defined only for dentals, and based upon the low and prominent peak of these consonants in /u/ context, clearly separated /u/ from /i,a/.

We found a positive correlation ( $r = 0.73$ ) between the values of the cue and the number of /u/ responses for each item. If we eliminated the /tu/ stimuli, the correlation jumped to 0.76.

#### 4.3. Classification of /a/

Despite the high identification rate of /a/ in the perceptual experiment, the subjects spontaneously told us that they could not actually identify the timbre of this vowel while they sometimes clearly identified the timbre of /i/ and /u/. Actually, their response could be considered as a double exclusion (of /i/ and /u/) rather than as an actual identification. We remarked that /a/ could be classified by the same process since most of /a/ items were characterized by low values of both cues. On the basis of this criterion, we can delimit an area containing all /ta,pa/ items but one, four /ka/, and only one non-/a/ item (one /pu/). The performance of listeners indicated that the identification of /a/ from /k/ (relatively high under training condition) could be improved.

### 5. CONCLUSION

We found interesting criteria for the identification of vocalic features from stop bursts. Using data from our corpus, the most intelligible contexts (/i/ from /t,k/, /u/ from /p,k/) as well as /a/ from /p,t/ have been well classified without specification of the consonant. Three contexts (/pi,ka,tu/), the less intelligible ones, characterized by a great amount of energy in mid-frequency range (2000-3500 Hz), were still confused one with the other. We were only concerned here with the more relevant acoustic patterns. Nevertheless, the performance of listeners convinces us that a better and a more complete identification of vocalic context from bursts could be obtained by the use of a conglomerate of acoustic cues.

### 6. REFERENCES

- [1] A. Bonneau. “Identification of vowel features from French stop bursts”, *Proc. ICSLP96*, vol 4, Philadelphia, PA, USA, 1996.
- [2] A. Bonneau, L. Djezzar, and Y. Laprie. “Perception of the place of articulation of French stop bursts”, *J. Acoust. Soc. Am.*, Vol 100(1), pp 555-564, 1996.
- [3] W. L. Cullinan and M. E. Tekieli. “Perception of vowel features in temporally-segmented noise portions of stop consonant CV syllables”, *J. Speech Hearing Res.*, Vol 22, pp 122-131, 1979.

<sup>1</sup>We considered that a context was identified as “V” by listeners if “V” responses were the most frequent.

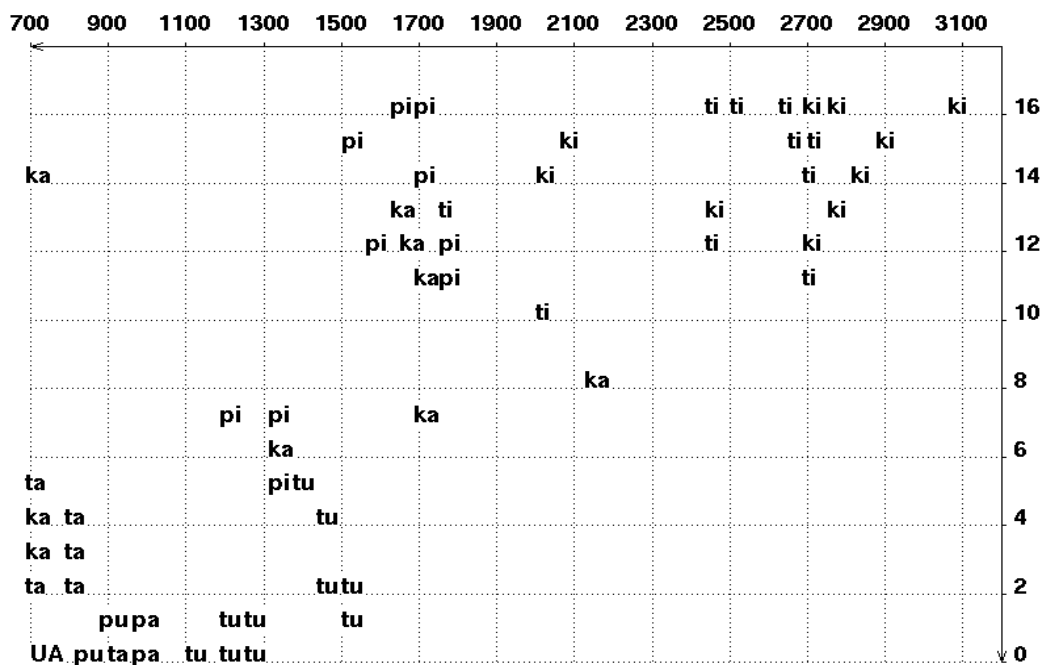


Figure 3: Low frequency limit of the frication noise, expressed in Hz, plotted as a function of the number of /i/ responses under the lack of training condition (16 responses per stimulus). See section 4.1 for a precise definition of the cue. The symbol UA at the bottom of the figure stands for 10 /ku/, 8 /pu/, 8 /pa/, 4 /ta/ and 1 /ka/ contexts.

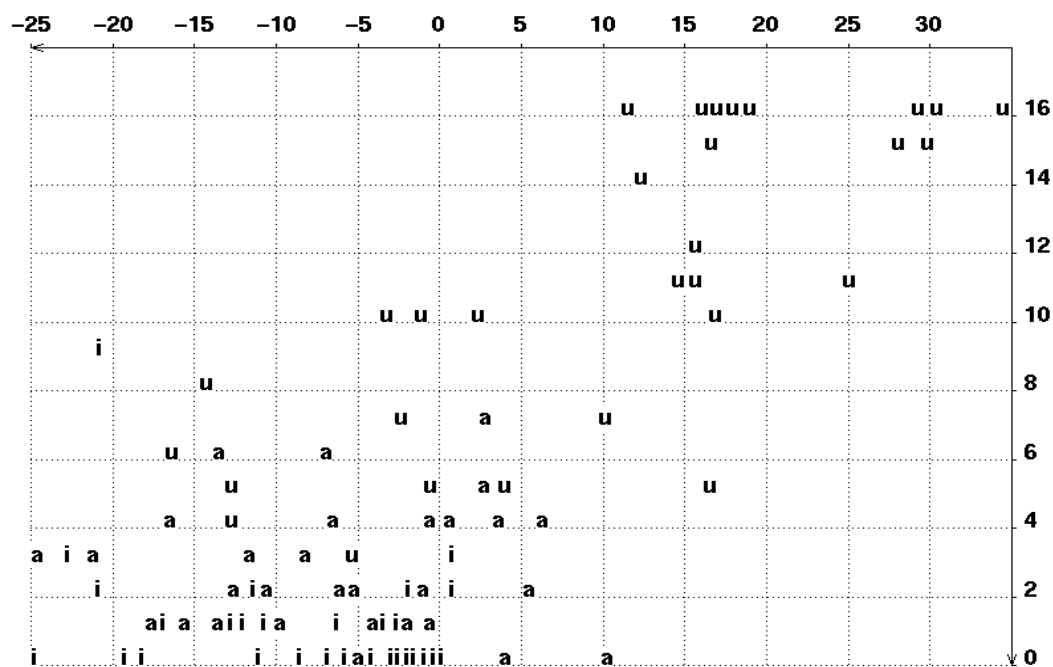


Figure 4: Emergence of a peak at the F2 frequency region of /u/ plotted as a function of the number of /u/ responses under lack of training condition (16 responses per stimulus). See section 4.2 for a precise definition of the cue. The values of the cue for all /pu,ku/ contexts but one are above 10.

Table 1: Center of gravity (CG) and low limit (Lim) of the burst calculated on the transient (Trans) and the fricative (Fric) segments of the burst. See section 4.1 for a definition of the low limit. The standard deviations are given in parentheses, data are expressed in Hz.

CV	CG(Trans)	CG(Fric)	Lim(Fric)
pi	<b>2523</b> (366)	<b>2883</b> (508)	<b>1537</b> (202)
ti	<b>3399</b> (308)	<b>4570</b> (533)	<b>2437</b> (315)
ki	<b>3601</b> (257)	<b>4037</b> (222)	<b>2612</b> (344)
pa	<b>2462</b> (383)	<b>1819</b> (466)	<b>737</b> (105)
ta	<b>3406</b> (384)	<b>3091</b> (496)	<b>718</b> (79)
ka	<b>3147</b> (206)	<b>3125</b> (337)	<b>1293</b> (556)
pu	<b>2230</b> (636)	<b>2149</b> (554)	<b>730</b> (66)
tu	<b>2772</b> (393)	<b>3591</b> (456)	<b>1337</b> (138)
ku	<b>2095</b> (677)	<b>1701</b> (353)	<b>700</b> (0)

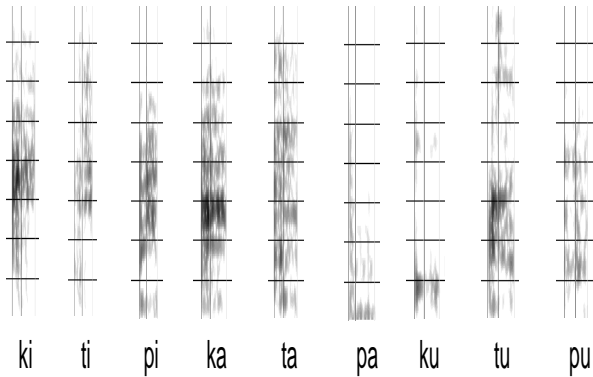


Figure 1: Example of a burst spectrogram for each CV context (the duration of the stimuli is 25 ms). The boundary within the spectrogram separated the transient from the fricative noise.

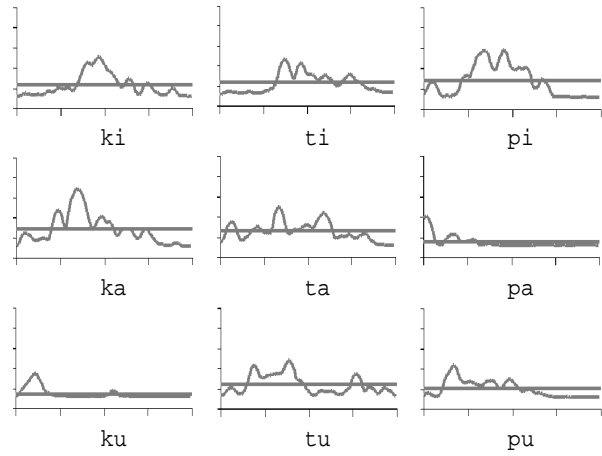


Figure 2: Example of a frication noise spectrum for each CV context. The vertical line represents the average energy of the whole spectrum. The corresponding spectrograms are displayed in Fig.1.

- [4] E. Fischer-Jørgensen. “Acoustic Analysis of stop consonants”, *Miscellanea Phonetica*, Vol 2, pp 42–59, 1954.
- [5] D. Fohr and Y. Laprie. “Snorri: an interactive tool for speech analysis”, *Proc EUROSPEECH89*, pp. 613–617, Paris, 1989.
- [6] B. Repp and H. B. Lin. “Acoustic properties and perception of stop consonant release transients”, *J. Acoust. Soc. Am.*, Vol 85(1), pp 379–396, 1989.
- [7] H. Winitz, M. E. Scheib, and J. A. Reeds, “Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech”, *J. Acoust. Soc. Am.*, Vol 51(4), pp 1309–1317, 1972.