# IDENTIFICATION OF VOWEL FEATURES FROM FRENCH STOP BURSTS

*Anne Bonneau*

CRIN-CNRS and INRIA-Lorraine
Bâtiment LORIA BP 239
54506 Vandœuvre-lès-Nancy
FRANCE
email: bonneau@loria.fr

## ABSTRACT

This paper deals with the perception of vowels from French stop bursts. The corpus was made up of 90 stimuli of 20-25 ms duration extracted from natural CVC and CV words. The syllables combined the initial stops /p,t,k/ with the vowels /i,a,u/. In order to cut off all traces of vocalic segment, bursts whose duration was too short were lengthened. Eight native speakers of French served as listeners in the experiment. Results showed that a burst onset which did not contain any traces of vocalic segment provided substantial vocalic information (the overall identification rate was 80%). The vowel /i/ was clearly identified from /t/ and /k/, and the vowel /u/ very clearly identified from /k/. The vowel /a/, with high identification rate, was often chosen in the absence of a clear vocalic timbre.

## 1. INTRODUCTION

The burst onset provides very reliable information about the place of articulation of stop consonants ([10, 11, 6, 7], among others). Knowledge about the identity of the following vowel, -whether its identity is revealed to the listeners, or whether the vowel, without any transition, is added to the burst signal-, leads to a slight improvement in stop identification [9, 1]. The aim of this paper is to investigate another kind of contextual information: the vocalic information provided by the burst itself.

Winitz et al. [12] reported that subjects could more than accidentally identify the adjacent vowels /i/, /a/ or /u/ from the entire burst of voiceless stops[1], the corresponding rates of recognition being 90%, 54% and 50%. Subjects were revealed the identity of the consonant during the vowel recognition task. Repp et al. [9] found that the vowel class, defined as the vowel and its nearest neighbours, could be identified by listening to whispered transients (61% while 33% represented the chance level). The vowels were not identified as accurately, but nevertheless better that could have been done by chance: 28% on the average, for the nine vowels of the corpus, with large individual differences since, from one listener to another, scores varied from 14% to 41%. The vowel /i/ was the most easily identified. Cullinan and Tekieli [3] found that vocalic cues were also present in the first milliseconds of natural bursts. They showed that listeners had more than by chance ability to identify the front-back feature 10 ms after the release, and the vowel height feature 20 ms after the release. When the burst lasted 30 ms, the front-back feature was correctly identified at least 90% of the time, but the height feature only 60% of the time. With 50% as the chance level, it appeared that only the front-back feature could be clearly identified by listeners. /I/ was the vowel best identified, /i/ being confused with it. The results of the three experiments thus concurred to show that the subjects could recover the high front characteristics of vowels /i/ from bursts. Ohde and Sharf[8] showed that vowel identification from bursts was greater in CV syllables than in VC syllables.

The aim of the present experiment was to verify whether the burst of French stops, containing no trace of vocalic segment, provides information about the identity of the following vowel.

## 2. PROCEDURES

We used the corpus and the stimuli of a previous experiment concerning the perception of the place of articulation of French stop bursts [1]. The corpus was made up of 90 isolated monosyllabic CV or CVC words, extracted from the French database BDSONS [2], which combined the three initial voiceless stops /p,t,k/ with the three vowels /i,a,u/. Each of the nine syllables was uttered twice by five male speakers.

Our stimuli were made up of fixed-length bursts of 20-25 ms duration containing no trace of vocalic segment. The duration was chosen in order to fulfill the following conditions: it must be sufficiently short to contain only syllable onset, and sufficiently long to allow the stimuli to be well perceived. In order to eliminate all traces of vocalic segment, burst signals whose duration was too short (three dental bursts and eleven labial bursts) were lengthened. For this purpose, we duplicated the burst tail end (essentially the weak frication noise of the bursts of /p/). Consequently, bursts longer than 25 ms were truncated, bursts shorter than 20 ms were lengthened to a final duration not exceeding 25 ms (the exact length of all the stimuli depending on the location of an appropriate zero crossing). We did not modify bursts with a duration of 20-25 ms. The preparation of the stimuli was done with the help of the Snorri signal editor designed by Y. Laprie [5] which displayed speech signals and spectrograms simultaneously and allowed to play back stimuli.

[1] Stop bursts can be decomposed in the following successive segments: the transient, the fricative segment which contains frication noise, and the aspirative segment [4]. The aspirative segment is generally absent from French bursts.

Eight native speakers of French (four women and four men) served as listeners in this experiment. They all reported to have normal hearing. None of them had participated in the previous experiment. The subjects listened to the sounds using Sennheiser HD520 II headphones, in a quiet room, with the sound volume adjusted to a comfortable level.

The listeners were first submitted to a training session containing various-length stimuli. The training corpus contained 27 syllables which combined the three voiceless stops with the three vowels /i,a,u/ and were uttered by three male speakers. These utterances were different from the utterances used in the test. From each of the 27 syllables, we generated four different stimuli: the burst plus 2/3, and 1/3 of the following vocalic segment, the burst segmented from the release to just before the onset of the first formant as well as fixed-length stimuli of 20-25 ms duration with no trace of vocalic segment. The training stage was conducted in three steps. In the first place, while reading the identity of the stimuli, subjects also listened to the training corpus. Next, they were tested on a randomized version of the corpus, and finally, while listening to the sounds, they compared their false and correct answers. The training lasted approximately 12 m.

The subjects then participated in three test sessions, each of them with a randomized version of the test corpus. Before the test sessions, they were asked to choose their response from among /i,a,u/ and to record it on answer sheets designed for the purpose -they put a cross in a column corresponding to their response. A ten minute break separated the training and the first session from the two remaining sessions. The stimuli were presented by blocks of nine tokens with the insertion of a ten-second inter-block interval and a four-second inter-stimulus interval. Each of the eight listeners gave 3 responses to each stimulus (one per session). We thus collected 24 responses per stimulus, and 2160 (24x90) responses on the whole.

## 3. RESULTS

The average identification rate was 80%. Tables I and II show the vowel confusion matrices for each context and each vowel. A two-way repeated analysis of variance, ANOVA, was performed to examine the effects of the vowel (three levels), and the stop (three levels). The Scheffé test was used for all post-hoc comparisons. The vowel effect was not significant ($p = 0.14$). Although the identification scores for each vowel were well distinct (84% for /a/, 80% for /i/, and 75% for /u/), the corresponding standard deviations were large. The consonant effect was significant ($p = 0.02$). The average identification scores were 88% from /k/, 79% from /p/ and 76% from /t/. The Scheffé test showed that the vowels were better identified from the consonant /k/. This result was expected since the articulation of /k/ varies a greatly deal with the articulation of the following vowel. The interaction between vowels and stops was very significant ($p < 0.0001$).

The subjects spontaneously told us that they could not actually recognize the timbre of the vowel /a/ while they sometimes clearly identified the timbre of the vowels /i/ and /u/. Paradoxically, the vowel /a/ obtained an high identification rate. In fact, it seems that listeners have chosen the /a/-response in the absence of a clear vocalic timbre. This probably explains the high rate of confusion of /i/ with /a/ (18%) and of /u/ with /a/ (24%). The opposite confusion rates of /a/ with /i/ and of /a/ with /u/ were lower (9% and 6%, respectively). It thus appears that the /a/-response has been given for the vowel /a/ and for the vowels /i/ and /u/ which were not clearly identifiable.

| | i | a | u |
|---|---|---|---|
| pi | **63** | 31 | 5 |
| ti | **90** | 9 | 1 |
| ki | **86** | 13 | 1 |
| pa | 1 | **85** | 14 |
| ta | 3 | **94** | 3 |
| ka | 21 | **74** | 5 |
| pu | 0 | 9 | **91** |
| tu | 2 | 61 | **37** |
| ku | 0 | 1 | **99** |

**Table 1:** Vowel confusion matrices. Scores (%) for each context. Columns display the responses.

| | i | a | u |
|---|---|---|---|
| i | **80** | 18 | 2 |
| a | 9 | **84** | 6 |
| u | 1 | 24 | **75** |

**Table 2:** Vowel confusion matrices. Scores (%) for each vowel. Columns display the responses.

Confusions between /i/ and /u/ were scarce (1% and 2%). With regard to the identification of vocalic features, these low confusion rates confirmed the very high intelligibility of the front-back feature, already noted in previous studies, especially in [3].

If we except the /a/-context, whose identification process was particular, we observed that the proximity of both the articulations of the stop and the following vowel constituted the best combination for the identification of vowels from bursts[2]. This was clearly shown by the identification rates of /u/ from /k/ (99%), of /i/ from /t/ (90%), and of /i/ from /k/ (86%). It is worth recalling that /k/ is velar when followed by the velar vowel /u/, and post-palatal when followed by the palatal vowel /i/. The high identification rates of the vowel /u/ from /k/ and of the vowel /i/ is in good agreement with results of previous experiments [9, 12, 3]. The vowel /u/ was very poorly identified from /t/, the places of articulation of the stop and of the following vowel being probably too far from each other to allow the vowel articulation to be achieved at the consonant release.

The vowel color seems to be, at least partially, determined by the global form of the spectrum. In /i/-context, especially for the den-

[2]This proximity also favoured the consonant identification (Bonneau *et al.*, 1996)

tal and velar consonants, we observed the presence of a long and intense frication noise, dominated by high frequency-peaks, which probably favoured the identification of the vowel (see Fig. 1). On the contrary, the concentration of energy in low frequencies probably favoured the identification of /u/ (see /u/ from /k/, 99%, and /u/ from /p/, 91%).

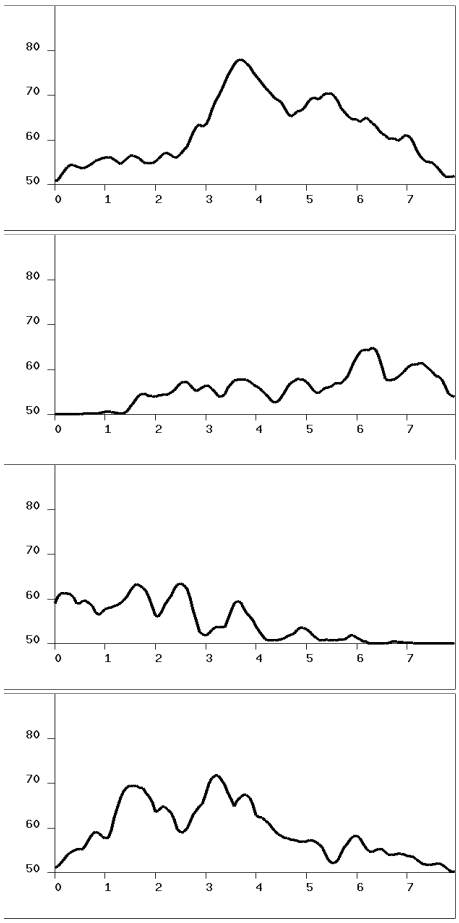


**Figure 1:** Spectra of the frication noise of the bursts (the transient is not included). From top to bottom: /k/ followed by /i/, /t/ followed by /i/, /t/ followed by /a/, /t/ followed by /u/. The frequency is expressed in kHz and the intensity in dB.

We investigated the relations between the results of our previous experiment [1], concerning the identification of stop bursts as well as the influence of the specification of the context on this identification, and the results of the present experiment. The same stimuli were used in both experiments. Table III displays the confusion matrices for the three stops /p,t,k/ in the three contexts /i,a,u/.

There was no clear relation between the identification of bursts (without specification of the context) and of vowels from bursts. It is worth noticing that most of the identification rates of both the stop and the vowel were concentrated at a small range of (relatively high) values, which was not favourable to find clear correlations between data. In fact, when all the contexts and all the data were considered, there was no correlation between vowel and burst identifications. Nevertheless, if 20% of the outliers were eliminated, there was only a weak correlation ($r = 0.6$). When we analysed the data of each syllable separately, we found only three correlations. There were negative correlations between the identifications of /i/ and /p/, and between the identifications of /i/ and /k/. We observed that the presence of very high frequency peaks on the bursts of /k/ and /p/ followed by /i/ led to a perfect identification of the vowel /i/ and to few /t/ responses. When the peaks were lower in frequency, the identification of the vowel decreased sligthly while that of /k/ and /p/ increased. We also observed a positive correlation between the identifications of /k/ and /a/. This correlation reminds us one of the main results of our previous experiment: the a priori knowledge about the identity of the following vowel /a/ led to an improvement of 18% in the identification of the consonant /k/. However, the results concerning the contexts should be confirmed with a larger corpus -there were ten stimuli per syllable.

| | p | t | k |
|---|---|---|---|
| pi | **89** | 6 | 5 |
| ti | 0 | **90** | 10 |
| ki | 0 | 28 | **72** |
| pa | **87** | 10 | 3 |
| ta | 15 | **79** | 6 |
| ka | 0 | 28 | **72** |
| pu | **85** | 6 | 9 |
| tu | 0 | **97** | 3 |
| ku | 0 | 2 | **98** |

**Table 3:** Stop bursts confusion matrices (results of a previous experiment [1]). Scores (%) for each context. Columns display the responses.

## 4. CONCLUSIONS

Our experiment showed that vocalic features could be identified from the bursts of French stops containing no vocalic segment. The front-back feature was particularly well identified since only 2% of the vowels /i/ and /u/ were confused one with another. The present study and our previous experiment concerning stop burst identification showed that the proximity between the articulations of the stop and of the following vowel always led to a clear identification of both the stop and the vowel.

## 5. REFERENCES

1. A. Bonneau, L. Djezzar, and Y. Laprie. Perception of the place of articulation of French stop bursts. *J. Acoust. Soc. Am.*, 1996. To be published.
2. R. Carré, R. Descout, J. Mariani, M. Eskenazi, and M. Rossi. The French language database: defining, planning, and recording a large database. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1984.
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.*, 22:122–131, 1979.

4. G. Fant. Stops in CV-syllables. Technical Report 4, Royal Institute of Technology, Stockholm, Sweden, 1969. Also published in Speech Sounds and Features, The MIT Press, Cambridge , Massachussets, and London, England, 1973.

5. D. Fohr and Y. Laprie. Snorri: an interactive tool for speech analysis. In *European Conference on Speech Communication and Technology*, pages 613–617, Paris, 1989.

6. D. Kewley-Port, D. B. Pisoni, and M. Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *J. Acoust. Soc. Am.*, 73(5):1779–1793, 1983.

7. D. Krull. Relating acoustic properties to perceptual responses: a study of Swedish voiced stops. *J. Acoust. Soc. Am.*, 88(6):2557–2570, 1990.

8. R. N. Ohde and D. J. Sharf. Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. *J. Speech Hearing Res.*, 20:543–554, 1977.

9. B. Repp and H. B. Lin. Acoustic properties and perception of stop consonant release transients. *J. Acoust. Soc. Am.*, 85(1):379–396, 1989.

10. K. N. Stevens and S. E. Blumstein. Invariant cues for place of articulation in stop consonants. *J.Acoust. Soc. Am.*, 64(5):1353–1368, 1978.

11. M. E. Tekieli and W. L. Cullinan. The perception of temporally segmented vowels and consonant-vowel syllables. *J. Speech Hearing Res.*, 22:103–121, 1979.

12. 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.*, 51(4):1309–1317, 1972.