EFFECT OF PRIOR KNOWLEDGE OF THE VOWEL ON THE PERCEPTION OF FRENCH STOP BURSTS

Linda Djezzar
CRIN-CNRS & INRIA Lorraine
B.P. 239 54506 Vandœuvre-lès-Nancy France

ABSTRACT

This paper presents a perceptual experiment on the effect of prior knowledge of the vowel on the identification of French stop bursts in natural speech. In order to evaluate the discrimination power of only spectral characteristics of the burst, stimuli consisted of fixed-length bursts of approximately 25 ms (neither VOT nor transitions are present). Results showed that knowing the identity of the following vowel caused a slight but statistically significant improvement of stop identification.

1 INTRODUCTION

Several studies have indicated that the release burst provides reliable information to correctly identify the stop place of articulation [1] [5] [6]. Nevertheless, in current recognition systems, the recognition rates of the palatovelars in front contexts and the dentals in rounded contexts are still far from human performance. Moreover, questions still remain about the role of the knowledge of the identity of the adjacent vowel on the identification of stops. In order to better understand the discrimination power of the burst for identifying the stop articulation place, three perceptual experiments were carried out: the first tested listeners' ability to identify the burst plus a segment of the following vowel, the second investigated the identification of fixed-length bursts independent of the knowledge of the following vowel, the third dealt with the effect of knowing the vowel on this identification. The first and the second experiments are described in [2] [3]. In the following section, we present them briefly in order to introduce the third experiment which is the central issue of this paper.

2 PERCEPTION OF STOP BURSTS WITHOUT KNOWING THE VOWEL

2.1 Preliminary experiment

We checked the listeners' ability to identify i) bursts plus a segment of the following vowel, and ii) bursts only. The results indicated that the presence of a vocalic segment, even a very short one, allowed an almost perfect stop identification. The high identification rates obtained for the burst-only stimuli encouraged us to propose the following experiment in which we tried to clarify the contribution of just the spectral burst characteristics.

2.2 Identification of fixed-length bursts

2.2.1 Corpus and stimuli

The stimuli were extracted from a corpus made of CVC and CV syllables, in which each of /p, t, k/ appeared in combination with each of the 11 vowels /i, e, y, ø, ɛ, a, œ, ɔ, u, o, ɔ̃/. The stimuli consisted of burst portions of the same duration (25 ms) with no remaining vocalic segment. Figure 1 presents the spectrogram of a few stimuli.

2.2.2 Major results

The average identification was 87%. Table 1 shows that, independent of context, all the identification rates were not only well above chance but also very high. A three way repeated analysis of variance, ANOVA, was performed in order to estimate the effects of the audition (2 auditions, one per session), of the vowel (8 vowels) and the stop (3 stops). Moreover, we used the Scheffé test for all post-hoc comparisons. These tests indicated significant main effects for all the parameters but only one significant interaction, the stop-vowel one.

To summarize, the burst onset provides reliable spectral information about the place of articulation of the stops, independent of explicit knowledge of the identity of the vowel. Neverthless, listeners' performance varied significantly depending on the following vowel and on the syllable. Does the knowledge about the identity of this vowel improve stop recognition, at least in the worst contexts? This question is the object of the following experiment.

Table 1. Consonant confusion matrice for fixed-length stimuli in eleven unknown vocalic contexts.

	p(%)	t(%)	k(%)
/pi/	89	6	5
/ti/	0	90	10
/ki/	0	28	72
/pe/	89	9	2
/te/	2	94	4
/pɛ/	87	11	2
/tɛ/	1	85	14
/kɛ/	1	28	71
/pa/	87	10	3
/ta/	15	79	6
/ka/	0	28	72
/py/	83	7	10
/ty/	0	72	28
/ky/	3	13	84
/pø/	97	2	1
/tø /	0	83	17
/kø /	2	6	92
/tœ/	0	98	2
/kœ/	1	12	87
/pu/	85	6	9
/tu/	0	97	3
/ku/	0	2	98
/po/	91	8	1
/to/	1	86	13
/ko/	0	1	99
/pɔ/	93	4	3
/tɔ/	8	83	9
/kɔ/	1	1	98
/pɔ̃/	92	6	2
/tɔ̃/	8	77	15
/kɔ̃/	2	0	298

3 EFFECT OF KNOWING THE IDENTITY OF THE VOWEL

Winitz *et al.* [8] have shown that, when listening to the entire burst of /p, t, k/, subjects could identify the adjacent vowel better than chance. According to Repp and Lin [6], the explicit knowledge of the following vowel slightly improved (3%) the consonant recognition. In order to verify whether Repp and Lin's results would be the same for natural (non whispered) bursts, we conducted an experiment to test the influence of the explicit knowledge of the vowel on the identification of the stops.

3.1 Identification of fixed-length bursts with knowing the vowel

3.1.1 Corpus and stimuli

The corpus was made up of 90 tokens: /p, t, k/ uttered twice by 5 male speakers, in each of the 3 vocalic contexts /i, a, u/. We used all the burst stimuli of the "vowel unknown" experiment appearing in /i, a, u/ contexts.

3.1.2 Subjects and procedure

The subjects were submitted to 3 sessions on 3 days, each day a session. The first session was a familiarization task, which consisted in hearing the stimuli of the preliminary experiment. The second session included 3 stages, one per vowel. At these stages, the identity of the vowel was revealed to the listener. Each stage was composed of a training phase, a test then a rest. The training phase was divided in 3 tasks.

- First, the subjects listened to the training corpus (/p, t, k/ appearing before each of /i, a, u/, and uttered twice by 3 male speakers) and simultaneously read the corresponding answers.
- Second, they underwent a test on the same corpus, but randomized, and they were asked to choose their response from among /p, t, k, ?/, the symbol /?/ means they were unable to supply an answer.
- Third, they compared their false responses by looking at the correct answers, and by simultaneously listening to the sounds.

The test task consisted in hearing the test corpus devised for one of /i, a, u/ (30 stimuli). The third session is a replication of the second one but the corpora were randomized differently. In this way, we obtained 2 auditions for each stimulus.

3.1.3 Acoustic analysis of the stimuli

In order to interpret the perceptual results, we analyzed the release burst of voiceless stops in all vocalic contexts.

- **Palatovelars.** As might be expected, the frequency of the most prominent peak varied as a function of both the place of articulation and the degree of rounding of the following vowel. More precisely, the mean values were 1020 Hz in the back context /u, o, ɔ, ɔ̃, ɑ̃/, 2500 Hz in the central context /a, ɛ̃, œ/ and 2800 Hz in the front context /y, ø, i, e, ɛ/.
- **Dentals.** In the rounded context, they had a prominent peak at a relatively low frequency, at approximately 2500 Hz. In the central context, the spectra of our dental stimuli were either flat or prominent at 1800 Hz (*Locus*).

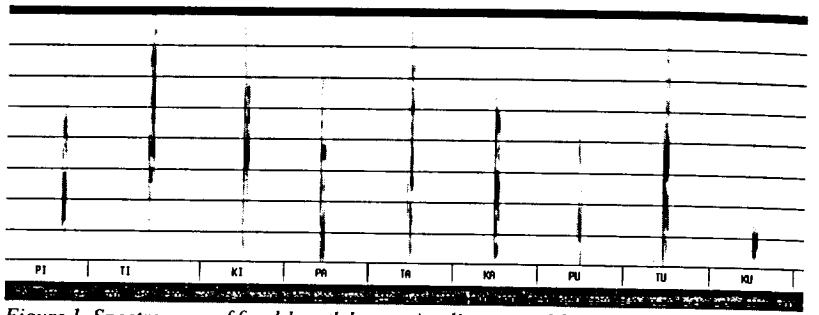

Figure 1. Spectrogram of fixed-length burst stimuli extracted from the syllables /pi, ti, ki, pa, ta, ka, pu, tu, ku/.

• **Labials.** The maximum was not situated in a well defined region, even for the same vocalic context. Nevertheless, the energy was stronger in low frequencies in the back context, and in mid frequencies in the central and front contexts.

Moreover, the global form of the spectrum did not always fit the templates proposed by Blumstein and Stevens [1]. In fact, the spectra of our dental stimuli followed by rounded vowels were not diffuse and those of our palatovelar stimuli followed by centrals and unrounded front vowels were not always compact [4].

3.1.4 Results

The overall identification rate was 89%, 4% higher than the identification rate obtained in the "vowel unknown" experiment for the same vocalic contexts /i, a, u/. A three way repeated analysis of variance was conducted in order to estimate the effects of the audition (2 auditions, one per session), of the vowel (3 vowels) and the stop (3 stops).

There was only one significant main effect, the vowel one [$F(2,166)=13.28$, p.c.<0.001], and one significant interaction [$F(4,166) = 7.67$, p.c.< 0.001], which occurred between the consonant and the vocalic context. According to the Scheffé test, /u/ was considered as the best vocalic context, and the interaction between /k/ and /i/ reduced the intelligibility of the consonant. Let us interpret briefly these results.

• **Context /u/.** In this context, the very high rate of identification of /k/ is most probably due to its particular spectral form (2 prominent peaks separated by a region without energy: generally the most prominent was situated at the F2 of the following vowel and the second one at 4000 Hz). Similar identification rates have been reported in [1] and [5]. Although the spectral form of /t/ was not diffuse, it was correctly identified; the spectral maximum frequency of /t/ in this context was clearly different of that of /p/ and /k/. /p/ was mostly confused with /k/ because its relatively prominent peak situated in the low frequencies.

Table 2. Consonant confusion matrice for fixed-length stimuli in three known vocalic contexts.

	p(%)	t(%)	k(%)
/pi/	92	1	7
/ti/	0	86	14
/ki/	0	22	78
/pa/	87	8	5
/ta/	8	79	13
/ka/	0	9	90
/pu/	93	3	4
/tu/	0	99	1
/ku/	1	1	98

• **Context /a/.** In most cases /k/ was confused with /t/ because: i) its non compact spectral form, and ii) the frequencies of its maxima were situated in the same region as those of /t/. /t/ was confused with /p/; the missclassified stimuli had a weak energy and their maxima were situated bellow 2500 Hz. /p/ was very well identified; it might be due to its weak energy.

• **Context /i/.** Most of the confusions were between /t/ and /k/. The latter was not always compact, and the frequencies of the maxima of the two stops were situated in the same region (3500 Hz-4500 Hz).

4 COMPARISON BETWEEN THE "VOWEL UNKNOWN" AND "VOWEL KNOWN" EXPERIMENTS

A four way repeated analysis of variance, was conducted in order to estimate the effects of the audition (2 auditions, one per session), of the vowel (3 vowels), of the stop (3 stops) and of the experiment (2 experiments).

The significant main effect of the experiment parameter [$F(1, 336) = 7.05$, p.c. < 0.01] proved that knowing the vowel improved the identification of stop bursts. There were significant effects of the vowel [$F(2,336)=28.83$, p.c. < 0.001] and a significant interaction between the consonant and the vocalic context [$F(4,336) = 14.40$, p.c. < 0.001]. The Scheffé test revealed that /u/ was the best vocalic context, and /a/ the worst one. The interaction between /k/ and /i/, /p/ and /u/, reduced the intelligibility of the consonant, while the interaction between /k/ and /u/ favored it. There was a marginal significant interaction between the experiment and the place of articulation [$F(2,336)=1.90$; p.c.< 0.25]. The Scheffé test indicated an improvement in the identification of /k/ in the "vowel known" experiment: there was no change for /k/ followed by /u/, already very well identified in the "vowel unknown" experiment, whereas the improvement for /k/ followed by /i/ and by /a/ was 6% and 18% respectively. The identification of /p/ was slightly improved (4%), while that of /t/ was reduced by 2%. Another more significant interaction was the triple interaction between the experiment, the stop, and the vowel [$F(4,336) = 2.42$, p.c. < 0.05]. The Scheffé test indicated that the improvement of the identification of /k/ followed by /a/ in the "vowel known" experiment was very significant while the improvement of /p/ followed by /u/ was only slightly significant. Consequently, vowel knowledge was especially beneficial to the identification of /p/ followed by /u/ and /k/ followed by /a/.

5 CONCLUDING REMARKS

The explicit knowledge of the vowel caused a slight but statistically significant improvement of stop identification. More precisely, knowing the vowel was of benefit to /k/ followed by /a/ and to /p/ followed by /u/. The increase appeared without an accompanying decrease of other stops in the same vocalic context, therefore the knowledge of the identity of the vowels /u/ and /a/ had only positive effects. On the other hand, the listeners' performance remained constant in the /i/ context. These deductions allowed us to appreciate how the vocalic information may help stop recognition. For this purpose, we are developing a system that uses vowel features to recognize the stop place of articulation.

ACKNOWLEDGMENT

These experiments were first reported in a Ph. D. thesis [3] supervised by J.-P. Haton, A. Bonneau and Y. Laprie. M. Depaix and F. Lonchamp provided precious advice for the succes of these experiments. The assistance of all these people is gratefully acknowledged.

REFERENCES

[1] Blumstein, S. and Stevens, K. (1979). Acoustic invariance in speech production : Evidence from measurements of the spectral characteristics of stop consonants. *J.A.S.A.*, 66(4):1001–1017.

[2] Bonneau, A., Djezzar, L. and Laprie, Y. (1993). Perception of French stop bursts, implications for stop identification. *Eurospeech*, volume 1, 693–696, Berlin.

[3] Djezzar, L. (1995). Contribution à l'étude acoustique et perceptive des occlusives du français. Ph. D. thesis, U. Henri Poincaré Nancy, France.

[4] Djezzar, L. (1995). Some new considerations about the spectral form of the stop bursts. To appear in *Eurospeech'95*, Madrid.

[5] Kewley-Port, D., Pisoni, D. and Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation. *J.A.S.A.*, 73(5):1779–1793.

[6] Repp, B. and Lin, H.-B. (1980). Acoustic properties ansd perception of stop consonant release transients. J.A.S.A., 85(1):379-396.

[7] Cullinan, W. and Tekieli, M. (1979). Perception of vowel features in temporally segmented noise portions of stop consonant CV syllables. *J. S. H. R.*, 22:103–12.

[8] Winitz, H., Scheib, M. E. and Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *J.A.S.A.*, 51(4):1309–1317.