VOICING ASSIMILATION AS A CUE FOR CLUSTER IDENTIFICATION

Christine Meunier
Laboratory of Experimental Psycholinguistics, University of Geneva,
Route de Drize, 9, 1227 Carouge, Switzerland
Tel. +41 22 705 97 41, FAX: +41 22 705 98 55, E-mail: meunier@fapse.unige.ch

ABSTRACT

It is well known now that speech chain is not constituted by discrete units. Speech sounds have an influence on other sounds directly in contact with them. We hypothesize that this influence is not noise but plays an important role for perception. An experiment is managed to evaluate the relative importance of two kinds of cues: those of phonetic distinctive features (voiced and unvoiced) and those of voicing assimilation (for liquids). Our results confirm that voicing assimilation of liquids plays an important role to identifie clusters: a/ the absence of assimilation cues increases reaction times; b/ subjects use assimilation cues in preference to distinctive features.

1. INTRODUCTION

One major problem for speech perception models is to understand how listeners can get linguistic information in spite of speech "enormous" variability. Some authors [1] consider that the perceptual system can use invariant patterns to identifie phonetic characteristics. At the opposite, others authors point out that variability could permit listeners to understand speech [2], [3], [4]. We subscribe to this second point of view, hypothesizing that without regular phonetic variability (coarticulation and assimilation) it could be difficult for listeners to identify phonemic information.
A previous study [5] showed a strong regularity of devoicing liquids within consonant clusters (unvoiced stop + liquid, /kl/ in "CLASSE"). These clusters are very frequent in French [6]. Three corpuses were used: clusters in isolated words, clusters in connected words and clusters separated by words juncture (the two consonants belong to two different words). Liquids are systematically and regularly affected by stops voicing characteristics. This analysis points out that the devoicing proportion depends on the liquid nature: /r/ is always completely devoiced, while /l/ is partially devoiced (only the first 35% of the liquid duration). The regularity of devoicing liquids lead us to hypothesize that voicing assimilation could be helpful for phoneme perception.
To check this hypothesis, we drew up an experiment where stop voicing cues and liquid voicing cues (by assimilation) are in conflict. We hope to show that assimilation cues are relevant for consonant cluster identification. We hypothesize that these cues are important for phoneme processing.

2. PROCEDURE

2.1. Stimuli elaboration

16 pairs of monosyllabic words were pronounced by a French native speaker (male, Swiss accent). Each pair was composed by words which start with a consonant cluster (stop + liquid). Both words of each pair differ by the voicing cue of the stop.
Two liquids (/l/ and /r/) and three pairs of stops were used (voiced: /b/, /d/, /g/, unvoiced /p/, /t/, /k/)
Examples:

- with /l/ **GL**ACE /glas/ - **CL**ASSE /klas/
- with /r/ **DR**AME /dram/ - **TR**AME /tram/

In these examples, the stop voicing cue is congruent with the acoustic cue of the liquids: when the stop is voiced, the liquids is completely voiced, and when the stop is unvoiced, the liquid is "devoiced" (see fig.5).
The recordings of original 16 stimuli constitute the first version (V1) of the test items.

A second version (V2) was obtained by manipulating the 16 original stimuli: the stops are acoustically separated from the rest of the word and cross-spliced in the two words of the pair.

Table 1: design of stimuli manipulation (see and listen to examples at the end of this paper, fig. 5).

	voiced stop	**unvoiced stop**
V1 originals	**gl[+v]as** [sound A0880S01.WAV]	**kl[-v]as** [sound A0880S02.WAV]
V2 manipulated	**g.l[-v]as** [sound A0880S03.WAV]	**k.l[+v]as** [sound A0880S04.WAV]

In this second version, the stop voicing cue is in conflict with the acoustic characteristics of the liquid:

- in /glas/ V2, /g/ is followed by devoiced /l/ (see fig.5).
- in /klas/ V2, /k/ is followed by /l/ voiced (see fig.5).

Thus, in V1 the voicing information for the both consonant in the cluster is redundant, while in V2 the stop voicing cue is in conflict with the acoustics characteristics of the liquid (/g/ is followed by /l/ partially devoiced, or /k/ is followed by /l/ entirely voiced).

In addition, fillers were recorded: 64 monosyllabic word starting either with a consonant cluster or a single consonant. The third of these fillers were manipulated.
For each subject, test items represent the third of all experimental items.

2.2. Experiment

40 subjects received an auditory stimulus followed by a visual presentation (a pair: GLACE - CLASSE). One word of the pair appeared right in a screen and the other left. The side of presentation was counterbalanced. They were asked to decide which word they heard by clicking as fast as possible right or left on a button box. Subjects Reaction Times and responses were analyzed. Reaction Times were measured from the beginning of the visual event. The ISI (Inter Stimulus Interval) was 1000 milliseconds.
Each subject heard half of test stimuli.

2.3. Hypotheses

2.3.1. Reaction Time

If the voicing cue of liquids is relevant for cluster perception, the contradictory information in V2 should disturb the subjects. Thus, we should observe an increase of reaction times (RT) for the manipulated version (V2).
Nevertheless, the increase of RT will mean that both cues (those of stops and those of liquids) are used by listeners. It means that RT will not increase for V2 if subjects only use one cue: the stop or the liquid voicing cue.

2.3.2. Responses

In this analysis we observe the competition between different cues: what is perceptually more relevant to identify the cluster? Three possible effects are expected:

- the stop voicing cue is perceptually more relevant (GLACE V2 will be identified as "GLACE")
  V1 and V2 will not differ
- the liquid voicing cue is perceptually more relevant (GLACE V2 will be identified as "CLASSE")
  V1 and V2 responses will be inverse
- liquid and stop voicing cues are as relevant as well (GLACE V2 will be identified sometimes as "CLASSE" and sometimes as "GLACE")
  V2 responses will be given at random

We suppose that these hypotheses should be moderated by some factors. We may observe differences according to the type of manipulation (voiced or unvoiced stops). Moreover, the importance of assimilation may play a role in cluster identification (/r/ devoicing is more important than one of /l/).

3. RESULTS

3.1. Time Reactions

The results of this experiment show an increase of reaction times when the voicing cue of the stop does not correspond to the voicing characteristics of the liquid: detection times are longer for V2 than for V1. This difference is significative by item ($p = .03$) and by subjects ($p = .0008$).

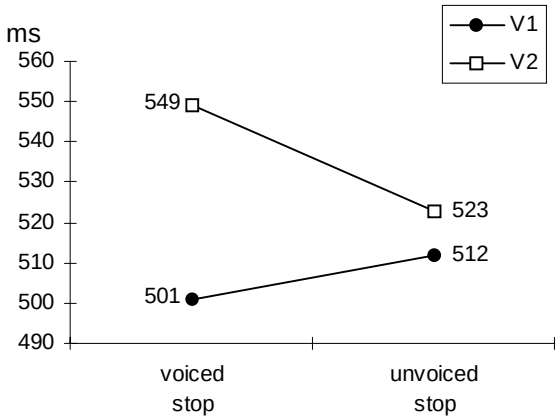

Figure 1: Detection latencies (in milliseconds) for the word detection.

But the difference between V1 and V2 is essentially due to the manipulated voiced stops (see fig. 1). This condition involve very slow RT if we compare them to those of the three other conditions. For voiced stops in V2, RT are significantly different from unvoiced stops in V2 ($p < .02$ by subject, but not significative by item) and from voiced stops in V1 ($p = .008$ by item, $p < .0001$ by subject). It seems that this condition (stop [+v], liquid [-v]) considerably disturbs the subject.
At the contrary, there is no significative difference between V1 and V2 for the unvoiced stop. This manipulated condition does not seem to disturb subjects. It would mean that there is no conflict between both cues of stops and liquids. However, these observations should be interpreted in the light of response analyses.

3.2. Responses

The response analyses show that subjects are considerably sensitive to the voicing cue of the liquids. Considering the manipulated stimuli, a great part of the responses (74%) are given with preference to the assimilation cue.

We observe different manipulation effects:

- unvoiced stops in V2 are identified as voiced (see fig. 2). The difference between V1 and V2 for unvoiced stops is clearly significative ($p < .0001$).
In this case, stop voicing cue does not play any role in cluster identification. The voicing information is only supported by the liquid.

• voiced stops in V2 are identified as unvoiced or voiced as well (see fig. 2). The difference between V1 and V2 for voiced stops is significative ($p < .0001$).
In this case stop and liquid voicing cues are relevant as well. Subjects seem to use both cues.

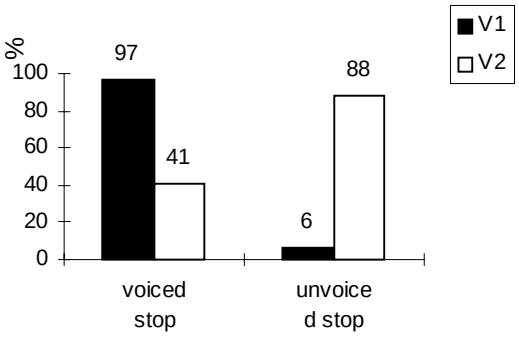

Figure 2: Word identification (in percentage) according to the stop voicing cue and the version type (V1 and V2).Only the "voiced" responses are reported here (each time a subject considered that the word started with a voiced stop).

We can explain these results by a phonetic explanation. Unvoiced stops are characterized by a burst but not by a closure (in word initial). In the other hand, voiced stop are characterized by a closure and a burst (see fig. 5). This means that unvoiced stop cues are very weak. Consequently, in V2 the competition between stop and liquid cues is really important when the stop is voiced. At the contrary, when the stop is unvoiced the competition is quite "unequal".
Response results are consistent with the reaction time analyses. Manipulated voiced stops show long RT because the conflict between stop and liquid voicing cue is really present. Manipulated unvoiced stops show short RT (as short as V1 ones) because the conflict between stop voicing cue is not relevant.

We observed above (in Reaction Times analyses) that there was no significative difference between V1 and V2 for the unvoiced stop. However, in the light of response analyses, we conclude that RT for unvoiced stops in V2 should be compared with those of voiced stops in V1 (as long as they are detected as voiced stops by subjects). The difference is significative by subject ($p=.0195$) but not by item.

Results are quite homogeneous regarding to different stops (see fig. 3). The unvoiced cue of /p/ seems to be more relevant (only 70% of manipulated /p/ are detected voiced) than the other (100% for /t/ and 96% for /k/). However, this observation should be moderated by the 10% of bad detection of /p/ in V1: this stop cue may be more ambiguous than the other. Moreover, the voicing cue of /g/ seems to be less relevant than the other in V2 (see fig.3). 68% of /g/ are identified as /k/ in the manipulated version. Consequently, the difference between /g/ and /k/ seems to be supported by the liquid voicing assimilation in clusters.

Regarding to liquid detailed results, we do not note important differences. We expected that the complete devoicing assimilation supported by /r/ could be perceptually more relevant than this of /l/ (only partial). Surprisingly, this is not the case: devoiced /r/ in V2 (the stop is voiced) involves 56% of "unvoiced responses" while devoiced /l/ in V2 involves 64% of "unvoiced responses" (see fig.4). This means that devoicing assimilation supported by /r/ is not more relevant than this of /l/.

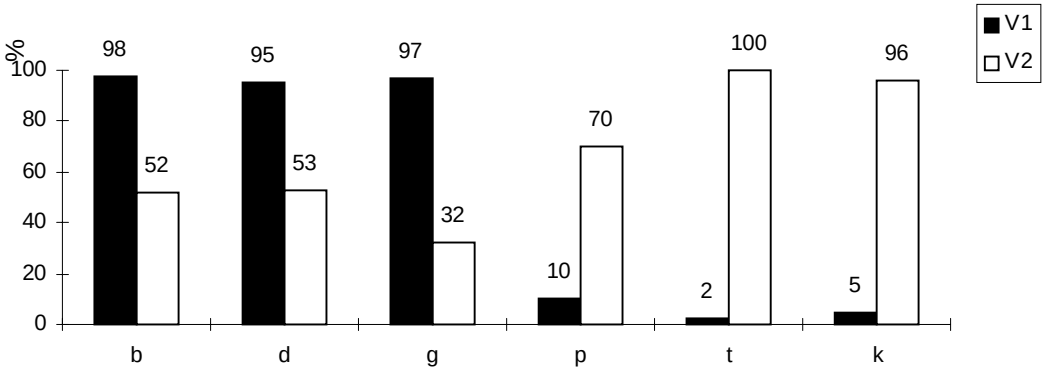

Figure 3: Word identifications (in percentage) in terms of stop type. Only the "voiced" responses are reported here.

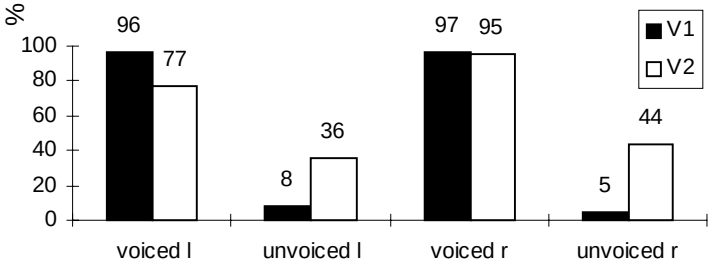

Figure 4: Word identifications (in percentage) in terms of liquid type. Only the "voiced" responses are reported here.

4. CONCLUSION

We hypothesized that assimilation cue should be relevant for cluster identification. Our experiment confirm this hypothesis. Subjects are faster to identify a word when stop and liquid voicing cues are congruent. It means that liquid assimilation plays a role in phoneme sequence processing. Moreover, to identify the cluster voicing cue, assimilation cues are preferably used than phonemic ones (for 74% of manipulated stimuli). Nevertheless, subjects responses are slow and unclear with manipulated voiced stops. This would mean that both cues (those of liquids and those of stops) are as relevant as well

In phonological terms, voicing is not a distinctive feature for liquids. Nevertheless, subjects seem to use voicing assimilation (for liquids) as a cue to determine the voicing characteristics of the preceding stop.

These results should help to understand the perceptual status of voicing assimilation in French and encourage us to go further in these investigations

5. REFERENCES

[1] Stevens K.N. and Blumstein S.E. (1978), "Invariant cues for place of articulation in stop consonants", in *J.A.S.A.*, 64, pp. 1358-1368.

[2] Elman J., Mc Clelland J. (1986), "Exploiting Lawfull Variability in the Speech Wave", in *Invariance and variability in speech processes*, Perkell and Klatt eds., M.I.T., London.

[3] Fowler and Smith, (1986), "Speech Perception as "Vector Analysis": An Approach to the Problem of Invariance and Segmentation", in *Invariance and variability in speech processes*, Perkell and Klatt eds., M.I.T., London.

[4] Rossi M. (1989), "De la quiddité des variables", *Actes du séminaire Variabilité et spécificité du locuteur: Etudes et Applications*, Marseille, éd. H. Méloni, pp. 11-31

[5] Meunier, C. (1994), *Les groupes de consonnes : problématique de la segmentation et variabilité acoustique*, Thèse de l'Université de Provence (Aix-Marseille I), Présentée le 7 mars 1994.

[6] Aubergé, V., Boë, L.J., Lefevre, J.P. (1988), "Lexique et groupes consonantiques", in *Proceedings of the 17th Journées d'Etudes sur la Parole*, Nancy, pp. 55-60.

Acknowledgments: I would like to give a special thank to Caroline Floccia for her helpful advice during the managing of this experiment.

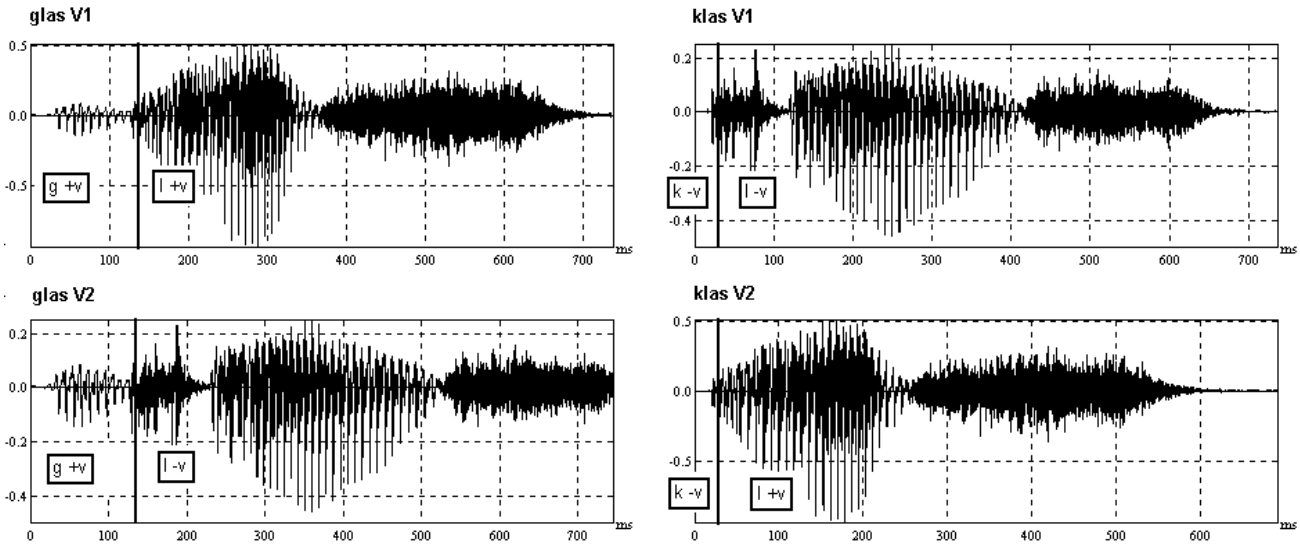

*Figure 5: example of pairs of stimuli in V1 and V2: the original voiced stop in "glas V1" **[sound A0880S01.WAV]**, the original unvoiced stop in "klas V1"**[sound A0880S02.WAV]**, the manipulated voiced stop in "glas V2" **[sound A0880S03.WAV]** and the manipulated unvoiced stop in"klas V2" **[sound A0880S04.WAV]**.*