RECOVERING MISSING PHONETIC INFORMATION FROM ALLOPHONIC VARIATION

Christine Meunier
Laboratoire Parole et Langage, CNRS ESA 6057, Aix-en-Provence, France

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. To test this hypothesis, a particular case of allophonic variation (liquid devoicing in unvoiced context) is tested. We predict that this variation could be used by listeners as far as it is robust cue. Two perceptual experiments (forced choice and phoneme monitoring) are managed to evaluate the role of allophonic variation for speech perception when expected phonetic information is missing. The results confirm our hypotheses that allophonic variations (devoiced liquids) are used as predominant cues by subjects to identify contextual phonemes.

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 identify phonetic characteristics. At the opposite, others authors point out that variability could permit listeners to understand speech [2], [3], [4, [5], [6]. We subscribe to this second point of view, hypothesizing that phonetic variability (coarticulation and assimilation) is expected by subjects, as long as this phenomena is frequent and regular.

A previous study [7] showed that liquids (/l/ and /r/) are systematically devoiced in unvoiced context as in /kl/ ("classe") or /pr/ ("presse"). These clusters (plosive + liquid) are the most frequent in French [8]. Consequently, this assimilation cue appears very frequently in speech production. To evaluate correctly this phenomena, three corpuses were analyzed: clusters in isolated words, clusters in connected words and clusters separated by words juncture (the two consonants belong to two different words). Results showed that liquids are systematically affected by plosives voicing characteristics. Morover, 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 frequency of liquids assimilation lead us to hypothesize that this allophonic information could be use by listeners through speech perception processing.

We consider devoicing of liquid as a robust allophonic cue, as long as it affects one of the distinctive feature of this phoneme (liquids are characterized like a voiced phoneme). Conversely, even if we suppose that liquid are affected by the presence of a voiced plosive, this allophonic cue does not strongly affect the canonical form of the liquid. As a consequence, we predict that listeners do not attribute the same perceptual weight to this two kinds of allophonic cue.

To check these hypotheses, we drew up two experiments in which the perceptual role of allophonic variation is tested. For both experiments two types of stimuli are constructed (original recordings of words, and manipulated version). In the first experiment subjects are expected to detect a word in a forced choice. We evaluate how they use allophonic variation to make their choice. In the second one, they are ask to detect a missing phonetic segment. Here, we predict that allophonic variation could be sufficient to recover the missing segment.

2. EXPERIMENT 1

2.1. Method

2.1.1.Stimuli. 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 (plosive + 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:

- voiced context with /l/: "**GL**ACE" /glas/
- unvoiced context with /l/: "**CL**ASSE" /klas/
- voiced context with /r/: "**DR**AME" /dram/ -
- unvoiced context with /l/: "**TR**AME" /tram/

In these examples, when the stop is voiced, the liquids is completely voiced, and when the stop is unvoiced, the liquid is "devoiced".

The recordings of 16 stimuli is the first version (V1) of experimental items. A second version (V2) is obtained by extracting the plosive from the carrier word (see Table 1). The first phoneme of the word is thus missing.

	voiced context	unvoiced context
V1 originals	gl[+v]as	kl[-v]as
V2 plosive extracted	(g) l[-v]as	(k) l[+v]as

Table 1: design of experimental stimuli.

Thus, in V1 the voicing information for both consonant in the cluster is redundant, while in V2 the stop voicing cue is only marked by the allophonic variation of the liquid. We will test if this variation is used (and to what extent) by listeners to recover stop missing information.

In addition, fillers were recorded: 64 monosyllabic word starting either with a consonant cluster or a single consonant (as in "vrai", "flic" or "basse", "terre", etc). The third of these fillers were manipulated (as in V2 experimental stimuli). Experimental stimuli represent the third of all stimuli. Whole stimuli were presented in random order.

2.1.2. subjects and procedure. 40 Swiss subjects (students of the University of Geneva) 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 were analyzed. Reaction times were measured from the beginning of the visual presentation. Each subject heard half of the test stimuli. The order of stimuli was counterbalanced across experiment.

2.2. Hypotheses

If listeners need the plosive acoustic features to identify the word, we suppose that the extraction would have great consequences on the lexical decision. Subjects may not be able to choose between the pair of words (/glas/ and /klas/) as long as the distinctive phoneme is not present. As a consequence, V2 results would show an important decrease of correct responses compared to V1 for voiced and unvoiced context as well. We should observe an increase of reaction times for V2.

On the other hand, if the plosive is absent, we can guess that subject would use other cues to distinguish both words. We suppose they will use the devoicing cue of the liquid as an information of the presence of an unvoiced phoneme. In this way, we hypothesize that subject will be disturbed by the absence of allophonic variation (voiced context) but will be less when allophonic variation is present (unvoiced context). Consequently we predict a weaker decrease of correct responses for V2 (compared to V1) in unvoiced context than in voiced context. Reaction times for V2 in voiced context should increase.

2.3. Results

2.3.1. Correct responses. We observe that word are correctly identified (from 79% to 99%) in original and in manipulated versions. Nevertheless, subjects do more errors for voiced context in V2 than in V1. But this is not the case for unvoiced context: subject could identify 94% of word when the plosive is absent (see Table 2 and Figure 1) as well as in original stimuli.

	V1 (original)	V2 (stop extracted)
Voiced context /glas/	99%	79%
Unvoiced context /klas/	97%	94%

Table 2: Percentage of correct responses for both types of stimuli (original and stop extracted) in both contexts (voiced and unvoiced).

2.3.2. Reaction times. We observe a clear difference between voiced and unvoiced contexts. In unvoiced context, reaction times do not differ significantly between V1 (originals) and V2 (extracted stops) ($F(1,30)=0.28$; $p=.6026$). Conversely, in voiced context, subjects are very slow to identify voiced context word in V2 ($F(1,30)=37.34$; $p<.0001$). The RT increase reflect the difficulty that have subject to choose between the two words proposed.

	V1 (original)	V2 (stop extracted)
Voiced context /glas/	510 ms *(49)*	816 ms *(194)*
Unvoiced context /klas/	537 ms *(61)*	552 ms *(91)*

Table 3: reaction times (in milliseconds) and standard deviations (in italic) for both types of stimuli (original and stop extracted) in both contexts (voiced and unvoiced).

These results confirm our hypothesis that subjects use unvoiced allophonic variation, when present, to make their decision. Conversely, when this information is absent (voiced context) subjects have difficulties to identify the correct stimuli. It seems that in unvoiced context subjects are not disturb by the absence of the plosive (no differences between response rates nor RT). Nevertheless, we do not know to what extent the absence of the plosive is not perceived. Subjects had to choose between to possibilities, thus the response is facilitated. To answer to this question, we managed a second experiment.

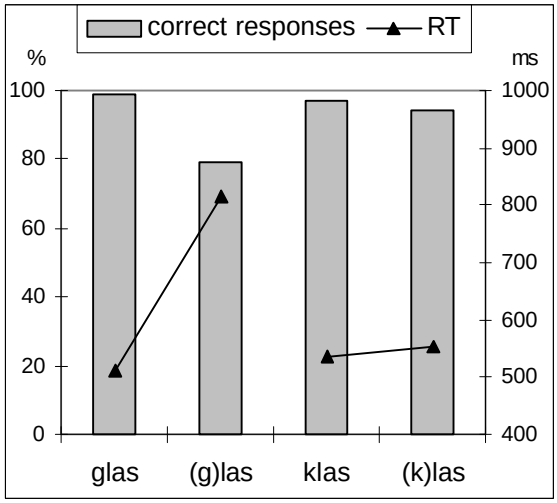

Figure 1: correct responses (left axis in percentage, gray bars) and reaction times (right axis in milliseconds, black points) in the forced choice experiment.

3. EXPERIMENT 2

3.1. Method

3.1.1. Stimuli. For this second experiment, we used exactly the same stimuli as in Experiment 1. V1 and V2 stimuli (see Table 1) were presented as experimental stimuli and the same fillers were added.

3.1.2. Subjects and procedure. 15 Swiss subjects (students of the University of Geneva) received first a visual target (plosive) and heard a carrier word. They were expected to detect, as fast as possible, the initial plosive in the carrier words. Each subject heard half of test stimuli. The order of stimuli was counterbalanced across experiment. Reaction times were measured from the beginning of the carrier words.

3.2. Hypotheses

In this experiment the task is different from this of Experiment 1. In the first experiment, subjects had to make a forced choice. In this way, we suppose that they could perceive the absence of the plosive but could answer even so. The forced choice could have helped them. They could identify the word using the allophonic information when it was present. In this second experiment subject had to detect an absent phoneme. So the specific role of allophonic variation is here clearly tested.

If the devoicing of liquids just help subject to decide which word is presented, but is not sufficient to recover plosive missing information, these plosives would never be detected in V2.

If the devoicing of liquid is a very strong cue, it may be sufficient to replace the presence of a plosive when absent. In this case, subjects would not note the absence of plosives and would detect them in unvoiced context for V2. V2 results may not be different from V1 in unvoiced context. Reaction times could inform us about the extent of plosive recoverability.

3.3. Results

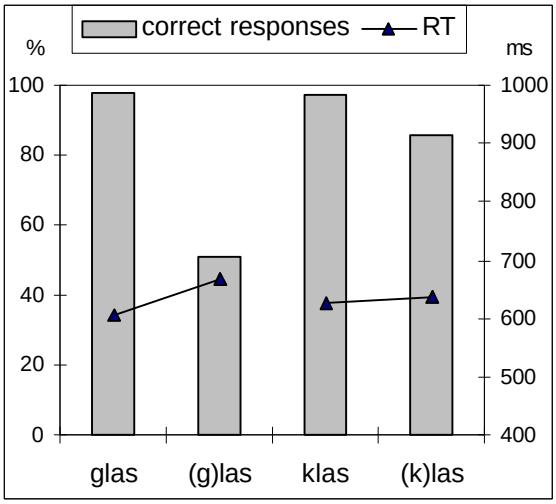

Figure 2: correct responses (left axis in percentage, gray bars) and reaction times (right axis in milliseconds, black points) in the phoneme detection experiment.

3.3.1. Correct responses. In the original stimuli (V1), plosives are completely detected in voiced as in unvoiced context (98% and 97%). When the voiced plosive is absent, we can observe that subjects can not detect the half part of them. At the contrary, the absence of the unvoiced plosive does not disturb the subjects in detecting them: 86% of plosives are correctly detected in spite of their absence.

	V1 (original)	V2 (stop extracted)
Voiced context /glas/	98%	51%
Unvoiced context /klas/	97%	86%

Table 4: Percentage of correct responses for both types of stimuli (original and stop extracted) in both contexts (voiced and unvoiced).

3.3.2. Reaction times. As in Experiment 1, we observe that, in unvoiced context, subjects do not respond slower in V2 than in V1 when they detect a plosive ($F(1,14)=0.24$; $p=.6284$). This means that they did not note that the plosive is absent. At the contrary, when a voiced plosive is identified, reaction times are slower in V2 than in V1 ($F(1,14)=6.28$; $p=.0252$).

	V1 (original)	V2 (stop extracted)
Voiced context /glas/	607 ms *(171)*	668 ms *(203)*
Unvoiced context /klas/	627 ms *(184)*	624 ms *(194)*

Table 5: reaction times (in milliseconds) and standard deviations (in italic) for both types of stimuli (original and stop extracted) in both contexts (voiced and unvoiced).

These results seem to confirm our second hypothesis: allophonic variation is a predominant cue to identify a unvoiced stop in liquid context. In unvoiced context, subjects do not note the missing plosive as long as phonetic information is completely carried by liquid allophonic information. Conversely, in voiced context, half of the plosives are not detected, and when they are, reaction times are slowed. This means that, here, subjects are disturbed by the missing information of plosive.

4. DISCUSSION

Both experiments confirmed our hypotheses that listeners use the allophonic information of the liquid to identify missing unvoiced plosives. Nevertheless, we should explain this results with additional facts. Unvoiced plosives are very poor in term of information quantity (no closure duration in initial, very short burst duration). In fact the duration of the unvoiced plosive is about 10 or 15 ms in initial position (only burst duration). This fact explain first that listeners look for other phonetic information around the plosive, and second, that the absence of the unvoiced plosive is not a strong lack of information. This is not the case for voiced plosive: the duration of the voiced closure informs listeners about its presence. In this way, its absence disturb subjects as long as this voiced closure is expected.

A previous study [9] showed that when subjects receive conflicting information (voiced plosive followed by a devoiced liquid, by cross-splicing manipulation), they can not decide what type of cluster they heard (/kl/ or /gl/). This seems to mean that both cues (voiced closure and devoiced) are used as well for the cluster identification.

The results confirm also that allophonic cues are used by listeners regarding to their frequency and to their acoustic importance. If the /l/ is also affected by allophonic variation in voiced context, this information is not sufficient to recover the missing phonetic information. At the contrary, liquid devoicing is a robust allophonic cue because of its acoustic weight (change of the distinctive feature) and its frequency. In a preceding study [7], we observed that allophonic variation differed in their frequency of apparition: voicing assimilation always affected liquids in clusters [plosive + liquid] or [fricative + liquid]. Conversely, plosive or fricative clusters (/bt/, /sv/., etc) showed very varying and heterogeneous voicing assimilation. We thus consider that liquids are marked by “variation ability”. This “no-resistance” to contextual acoustic characteristics lead us to hypothesize that liquids, as other phonemes, carry contextual information. As a consequence, this “variation ability” should not be considered as noise but as “syntagmatic” information, in other words information on preceding or following phonemes.

We hypothesized that allophonic information should be relevant to recover phonetic missing information. Our experiments confirm this hypothesis. Subjects do not note the absence of unvoiced plosives because of allophonic information.

In phonological terms, voicing is not a distinctive feature for liquids in French. Nevertheless, subjects seem to use voicing assimilation (for liquids) as a cue to recover the voicing characteristics of the preceding plosive. Even if listeners are not conscious of the presence of allophones, they use their informational cues to access phonemic characteristics.

These results help us to understand the perceptual status of voicing assimilation in French and this of phonetic variation in general. We hypothesize that phonetic units are different regarding to their resistance to variations. The more phonetic units are "variable", the more they inform listeners on contextual characteristics.

ACKNOWLEDGMENTS

This research was made possible by grants from the F.N.R.S. (11-39553.93). .I would like to give a special thank to Uli Frauenfelder and Caroline Floccia for their helpful advices during the managing of these experiments.

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] Repp, B.H. (1982) "Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception", *Psychological Bulletin*, 92, 1, 81-110.

[6 Labov W. (1986), "Sources of Inherent Variation in the Speech Process" in *Invariance and variability in speech processes*, Perkell and Klatt eds., M.I.T., London.

[7] 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.

[8] 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.

[9] Meunier C. (1997) "Voicing assimilation as a cue for cluster identification", *Proceedings of the Fifth European Conference on Speech Communication and Technology*, Rhodes, CDROM.