HOW ACOUSTIC ANALYSES CAN IMPROVE SEGMENTATION CRITERIA

Christine Meunier

Laboratory of Psycholinguistic, 9 route de Drize, CH. 1227 Carouge, Switzerland

ABSTRACT

This paper shows how the acoustic analysis of transition phases between two phonetic units can improve the criteria of speech segmentation. We propose a system of rules based on acoustic analysis (of transition phase), phonetic distribution (type of cluster), and syntaxic context (isolated word, word juncture, etc). These rules allow us to validate segmentation criteria and to make them more precise.

1. INTRODUCTION

The elaboration of segmentation criteria and labelling conventions raises several types of problems [1]. One major problem is the lack of constancy in the segmentation criteria due to the of variability of speech. Indeed, we observe different transition phases between the same two consonants of a cluster (figures 2 and 3).

In addition, segmentation operations presuppose having prior criteria and, at the same time, those criteria can only be elaborated through the regular exercice of segmentation. A neophyte segmentator in this paradoxical situation experiences a short period of instability during which the criteria are established. Once they are stable, a human expert can describe the qualitative nature of the boundaries but he only has an approximative idea of the quantitative representation of boundaries.

Regarding the problem raised by criteria elaboration, we decided to divide the analysis into three different steps. First, basic criteria were drawn up; this means that only qualitative observations were made. Second, the analysis of the labelling data produced a quantitative representation of consonant transitions within clusters. Third, quantitative results are compared to qualitative observations.

2. METHODOLOGY

Some basic methodological options of our general study of French consonant clusters [2] are presented here to make this specific procedure clear.

2.1. Consonants

We classified the French consonants relating to articulation manner:

- "S" are Stops: p t k b d g
- "F" are Fricatives: f s S v z J
- "V" are Vocalic consonants: l r m n j w

Consonant clusters are the logical combinations of the 3 consonant classes:

- SS (as in "obtu")
- FF (as in "asphalte")
- VV (as in "parmi" or in "lui")
- SF (as in "psychologue")
- SV (as in "bleu")
- FV (as in "flou")
- FS (as in "style")
- VS (as in "halte")
- VF (as in "farce")

2.2. Segmentation principles

The segmentation we practiced is based on a methodology developped by Autesserre and Rossi [3]. It is an accurate segmentation and hierarchically organized labelling: phonetic units are segmented at three different levels: *macro-classes* (classes of phonemes), *phases* (onset, stable part and release of the phonetic units) and *attributes* (acoustic events). This segmentation procedure produces sub-segments which can be used to locate the boundary between two consonants. The acoustic analysis of these sub-segments will give us the statistical representation of the boundaries within clusters and also will help to quantify segmentation criteria.

2.3. Acoustic analysis principles

In segmenting and labelling a speech database, we were led to adopt a specific acoustic analysis methodology: we distinguished segmentation specific features (articulation manner and voicing) from coarticulation features (articulation place) [2].

An accurate analysis of the acoustic realisations of consonant clusters allows us to draw up a list of transition phases.

We define the sub-segment in the following way: it is a transitory segment which appears at and around the boundary between the two consonants. The transition phase between C1 and C2 is realised either directly (<u>Direct Passage</u>: the acoustic characteristics of C2 directly follow those of C1), or by a <u>Transitory Segment</u> which affects one of the two consonants. The transitory segment supposes a contrast to the acoustic characteristics of the consonants.

Consonant characteristics are described as follows:

-V: vocalic formant structure
-F: voiced or unvoiced noise
-S: silence or voicing + burst

Four types of transitory segments are considered:

- <u>vocalisation</u>: insertion of a vocalic or a vowel unit
- <u>consonantisation</u>: insertion of noise
- <u>devoicing</u>: disappearance of voicing
- <u>voicing</u>: appearance of unexpected voicing

2.4. Basic segmentation criteria: observation of the transition phase

A set of basic qualitative criteria are drawn up. They will be compared to the statistic analysis results. In order to give an impression of the nature of these criteria, we have presented below the transition phases for each cluster class.

DP is Direct Passage and TS is Transitory Segment:

-SS: DP or TS: devoicing
-FF: DP or TS: devoicing
-VV: DP
-SF: DP or TS: devoicing
-SV: if F voiced, DP
if F unvoiced, DP or TS: devoicing
-FV: if F voiced, DP
if F unvoiced, DP or TS: devoicing
-FS: DP or TS: devoicing
-VS: if F voiced, DP
if F unvoiced, DP or TS:
devoicing or consonantisation
-VF: if F voiced, DP
if F unvoiced, DP or TS:
devoicing or consonantisation

As we can observe, there are numerous types of transition phase between two consonants for each consonant cluster class. Only a quantitative analysis of the boundaries can make more precise and improve the segmentation criteria.

3. LINGUISTIC MATERIAL

It is important to base the acoustic analysis upon samples taken from different speech contexts. Indeed, we have observed that different speech types lead to specific types of transition phases within clusters. Therefore, our segmentation and analyses are based on three different speech contexts.

3.1. Isolated words

Our first segmentation criteria were elaborated from a corpus (Clusters: ACC01 to ACC05) of the BDSONS (the French sounds database [4]). This corpus is composed of 5 lists of 41 isolated mono or bi-syllabic words containing clusters. 12 speakers (6 male and 6 female) read them. The clusters were representative of their distribution in French.

3.2. Integrated words

The second corpus was based upon words integrated in the same sentence: "Ce n'est pas ___ qu'il faut dire"..

These sentences were read twice by two speakers.

3.3. Words in juncture

This corpus was made up of sentences in which clusters crossed word boundaries.These sentences were read twice by two speakers.

In this corpus we considered two levels of junctures: the first a <u>major boundary</u> and the second a <u>minor boundary</u>. In fact, the sentences were of the very simple syntactic structure: NP+VP. The first type of juncture was between NP and VP (the major syntactic boundary), the second one was inside VP (between V and N, the minor boundary). We expected to obtain different acoustic effects as a function of the type of juncture which separated the first and second consonant (C1 and C2).

4. RESULTS AND RULES

4.1. The distribution of clusters transition phases

The results of our acoustic analysis allows us to distinguish two different classes: clusters including a vocalic consonant, and clusters without vocalic consonant. Vocalic consonants are systematically and regularly affected by assimilation effects [2] [5]. Acoustic variability obeys specific rules when a vocalic consonant is present in the cluster. Assimilation effects are also present in clusters without vocalic consonant, but the phenomena are quite irregular and do not follow specific rules.

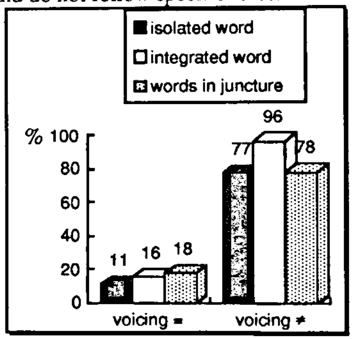

Figure 1: percentage of transitory segments within the three corpuses. Voicing similarity (voicing =) or opposition (voicing ≠) between the both consonants of the clusters are presented.

Voicing opposition is the factor which leads to the most important number of Transitory Segments (figure 1). Also the most frequent Transitory Segment is devoicing.

4.2. Rules

Rules are deduced from the acoustic analysis. The hierarchical presentation of rules follows the decreasing importance of variation in the transition phases of the clusters. The nature of the transition phase and their frequency are given (abbreviations are explained above).

1: If there is voicing opposition within cluster: 2 else TS0
2: If V is in cluster: 3 else 11
3: If cluster is pronounced in isolated word: 4 else 9
4: If V is C2: 5 else 8
5: If cluster is SV: 6 else 7
6: If V is /r/: **TS1** else **TS2**
7: If V is /r/: **TS1** else **TS3**
8: If V is /r/: **TS4** else **TS5**
9: If V is C2: **TS1** else 10
10: If V is /r/: **TS6** else **TS7**
11: If C1 is voiced: 12 else 13
12: If there is a minor boundary between C1 and C2: **TS8** else **TS9**
13: If cluster is FF: **TS10** else 14
14: If there is a minor boundary between C1 and C2: **TS11** else **TS12**

Lists and description of each Transitory Segment (**TS**):

TS0: irrelevant TS
TS1: TS = 100% of devoicing C2
TS2: TS > 90% of devoicing C2
TS3: TS > 60% of devoicing C2
TS4: TS > 5% of consonantisation C1
TS > 90% of devoicing C1
TS5: TS > 20% of consonantisation C1
TS > 15% of devoicing C1
TS6: TS > 30% of consonantisation C1
TS7: TS > 25% of consonantisation C1
TS > 20% of devoicing C1
TS8: TS = 100% of devoicing C1 (SS, FF, FS)
TS > 50% of devoicing C1 (SF)
TS9: TS = 100% of devoicing C1 (SS, FS)
TS > 50% of devoicing C1 (SF)
TS > 30% of devoicing C1 (FF)
TS > 50% of vocalisation C1(FF)
TS10: TS = 100% of devoicing C2
TS11: TS > 50% of devoicing C2
TS > 50% of voicing C1
TS12: TS > 80% of devoicing C2 (SS, FF)
TS > 30% of devoicing C2 (SF, FS)
TS > 30% of voicing C1 (SF, FS)
TS > 30% of vocalisation C1 (SF, FS)

5. DISCUSSION

These rules show that devoicing of a vocalic consonant in C2 position is the most regular phenomenon (TS1, TS2, TS3). This phenomenon becomes quite rare when the vocalic consonant is in C1 position (TS4, TS5, TS6, TS7), except for /r/ which is the most assimilated phonetic unit [2]. Tendencies are more uncertain for clusters composed of stops and/or fricatives. The devoicing of voiced C1 is frequent if C2 is unvoiced (TS8, TS9), but it depends on the cluster types.

Devoicing is the most important transitory segment. Nevertheless, the devoicing tends to be more progressive when a vocalic consonant is present in the cluster, and more regressive when clusters are made up of stops or fricatives.

5.1. Qualitative and quantitative segmentation criteria

These rules and, more precisely, the acoustic analysis of consonant clusters, confirm the basic segmentation criteria elaborated above. They bring quantitative precision for qualitative criteria. For SV and FV the devoicing is confirmed and represents the quasi systematic type of boundary; for VV, the Direct Passage is the most common boundary.; for FS devoicing is as frequent as voicing; etc.

Knowledge of the quantitative distribution of consonant cluster transition phases helps us to elaborate more robust segmentation criteria.

6. CONCLUSION

It is quite difficult to elaborate robust segmentation criteria. Nevertheless a human expert can formalize qualitative information and has a quite good representation of quantitative information. An accurate segmentation associated with an analysis of transition phases within clusters can make up for the lack of quantitative descriptions.

This additional segmentation tool should be useful for different speech technologies: Automatic Segmentation (improvement of reliability and precision), Speech Synthesis (improvement of clusters synthesis), Speech Recognition (improvement of segments and sub-segments identification).

REFERENCES

[1] Abry, C., Autesserre, D., Barrerra, C., Benoît, C., Boë, L.J., Caelen, J., Caelen-Haumont, G., Rossi, M., Sock, R., Vigouroux, N. (1985), "Propositions pour la segmentation et l'étiquetage d'une base de données des sons du français", *Actes des 14èmes Journées d'Etude sur la Parole*, Paris, pp. 156-163.
[2] 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.
[3] Autesserre, D., Rossi, M. (1989), "Une méthode de segmentation et d'étiquetage des groupes consonantiques de la base de données des sons du français", *Journal d'Acoustique*, 2, pp. 311-322.
[4] Descout, R., Sérignat, J.F., Cervantes, O., Carré, R. (1986), "BDSONS: une base de données des sons du français", *12ème Congrès International d'Acoustique*, A4-7, Toronto.
[5] Meunier C. (1994) "Les facteurs responsables de l'assimilation: analyse de trois types de groupes de consonnes et de leur miroir", *Actes des 20èmes Journées d'Etude sur la Parole (JEP)*, Lannion, pp 447-452.

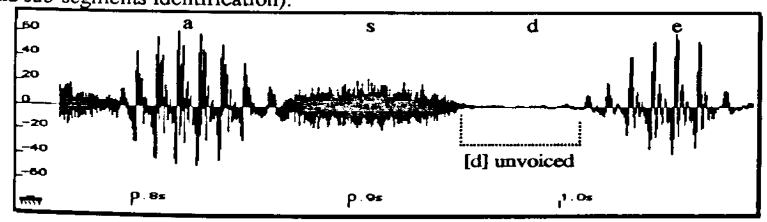

Figure 2: cluster /sd/ in the sentence "Les promeneurs cha<u>ssent d</u>es papillons". The speaker produced a progressive assimilation: [d] is clearly unvoiced.

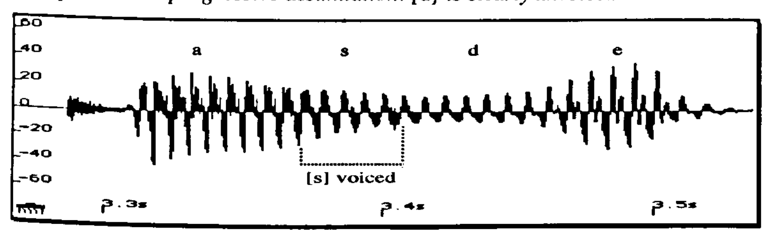

Figure 3: cluster /sd/ in the sentence "Les promeneurs cha<u>ssent d</u>es papillons". The speaker produced a regressive assimilation: [s] is clearly voiced.