Do different boundary types induce subtle acoustic cues to which French listeners are sensitive?

Odile Bagou¹, Sophie Dufour¹, Cécile Fougeron², Alain Content³, Ulrich H. Frauenfelder¹

¹ Lab. de Psycholinguistique Expérimentale, Université de Genève, Suisse
² Lab. de Phonétique et Phonologie (UMR 7018) CNRS/Sorbonne Nouvelle, Paris, France
³ Laboratoire de Psychologie Expérimentale, U. Libre de Bruxelles

Odile.Bagou@pse.unige.ch, Sophie.Dufour@pse.unige.ch, cecile.fougeron@univ-paris3.fr, acontent@ulb.ac.be, Ulrich.Frauenfelder@pse.unige.ch

Abstract

This paper examines the production of perception of three types of phonological boundaries. In the first part, we extended our previous acoustic analysis to confirm that French speakers mark word and syllables boundaries differently in enchaînement sequences. The durational properties of vowels and consonants were compared in 3 boundary conditions: (A) enchaînement (V1#CV2), (B) word-initial consonant (V1#CV2), (C) syllable onset consonant (V1.CV2). Results showed that the three boundary conditions are varying in subtle durational differences on V1 and C. In the second part, the sensitivity of French listeners to these acoustic cues was evaluated. Preliminary results showed that participants are sensitive to durational differences, at least for discriminating between syllable and word boundaries. Implications of these results for lexical segmentation are discussed.

1. Introduction

According to Content, Kearns and Frauenfelder [1], French listeners take syllable onsets to be possible word onsets. According to their Syllable Onset Segmentation Heuristic (SOSH) [2], French listeners use syllable onsets as alignment points with the mental representations stored in the lexicon. However, the efficiency of SOSH may be called into question by instances in which syllable onsets do not correspond to word onsets. Indeed, several phonolological phenomena are traditionally described as processes involving the complete resyllabification of a word-final consonant to the onset of the following syllable across word boundaries [3]. Resyllabified consonants have been described as phonologically identical to word-initial consonant. In French, three types of phonological processes require “so-called” resyllabification: (1) elision, where two consonants are linked due to the deletion of a schwa (e.g. /pa/ko/ “pas d’role”); (2) liaison, where a final latent consonant is produced and linked to the following initial vowel (e.g. /deza/ “des amis”, ‘friends’); and (3) enchaînement, where a word-final consonant is resyllabified in the surface form with the following word-initial vowel (e.g. /belam/, “bel ami”, ‘nice friend’). These phenomena constitute a challenge for SOSH since they should lead it to activate misaligned lexical candidates and slow or prevent the recognition of the intended word. The efficiency of this heuristic is thus reduced when an underlyingly word-final coda consonant surfaces as the onset of the following word, that is, when an apparent misalignment emerges between syllable and word boundaries. It is therefore important to determine whether the contrast between a resyllabified sequence and a sequence with an underlying onset is neutralized or whether phonetic cues distinguish the two.

Several psycholinguistic studies have shown that resyllabified consonants involved in elision [4], enchaînement [5], or liaison [6] differ acoustically from underlying word-initial consonants. For example, Dumay et al. [5] showed that both V1 and C durations are longer in cases of enchaînement (e.g. tante roularde ‘crafty aunt’), compared to sequences including an underlying word-initial cluster (e.g. temps troublants ‘disturbing weather’). More interestingly, Dumay et al. [5] showed that listeners can use these cues in on-line segmentation processes. In particular, they showed that the spotting of target-words is delayed when they are not aligned with syllable onsets (e.g. “tante” in /ta/ma/ extracted from “temps troublant”). In the light of these results, we hypothesize that SOSH could be guided by acoustic cues to avoid missegmentations.

This paper is divided into two parts. First, along the lines of Dumay’s work, we extended our previous analysis of the overall durational data of 8 speakers [6; 7]. We examine the cues to syllable and word boundaries by looking at cases of enchaînement between a single consonant and a vowel (V1#CV2). A third condition consisting of a sequence containing a syllable onset (not at word-onset) was tested, which allows for a three-way comparison with sequences containing a so-called resyllabified consonant and sequences containing a word-initial consonant. This non-word-initial syllable onset condition makes it possible to test for a phonetic difference between word and syllable boundaries. Then, we expose the specific acoustic characteristics of the speaker used in the perceptual experiments. In the second part, we present two preliminary perceptual experiments aiming at testing the ability of native French listeners to use these acoustic cues.

2. Production data

2.1. Corpus construction

As part of a larger project looking at speech segmentation and particularly the plausibility of SOSH [8], we created a database of 91 triplets varying with respect to the boundary condition (Table 1). For each triplet, the same critical V1CV2
sequence differed in the boundary location (before or after C) and the type of boundary involved in the sequence (syllabic or lexical).

Table 1: Boundary conditions studied

<table>
<thead>
<tr>
<th>Boundary Conditions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Enchaînement</td>
<td>V1,C#V2</td>
</tr>
<tr>
<td>B Word-boundary</td>
<td>V1#CV2</td>
</tr>
<tr>
<td>C Syllable-boundary</td>
<td>V1.CV2</td>
</tr>
</tbody>
</table>

Lexical items were extracted from the Brulex database [10] by selecting words ending (étape ‘stage’) and beginning (peureux ‘fearful’) with a same consonant (/p/). One of the major constraints was to choose items that could be transformed into other French words by removing the the consonant (étape ‘stage’ > état ‘state’; peureux ‘fearful’ > heureux ‘happy’). Then, 91 pairs of words were constructed in order to obtain “Noun+Adj.” or “Adj+Noun” sequences (e.g. un état peureux ‘a fearful state’ vs. une étape heureuse ‘a happy stage’; pleine aisance ‘full affluence’ vs. plate naissante ‘nascent wound’). For the third condition, it was unfortunately not possible to find enough French words with the required structure. Therefore, 91 non-words with the same VCV sequence were created.

Each member of the triplet was included in a syntactically and semantically plausible carrier sentence that controlled for length, sequence position in sentence, and prosodic structure. Particular care was taken to place the two-word sequences in a single accentual phrase to avoid a final accent on the first word. The carrier sentences were then randomized and recorded in a sound-attenuated booth by 8 Swiss French naive speakers.

The V1CV2 sequences were then excised from the carrier sentences and durational analysis were made with Praat [7]. The productions of the 8 speakers were assigned to three phoneticians who independently segmented the critical sequences. One speaker was analysed by the three phoneticians to ensure the validity of the measurements. A high positive correlation between the acoustic measurements of the three labellers was found (between .8 and .9).

2.2. Durational analysis

The 35 best triplets were selected on the basis of a pre-test that evaluated the syntactic and semantic plausibility of the carrier sentences. To test for differences in duration between sequences with enchainement (condition A) and sequences involving lexical or syllabic onsets (conditions B and C), an analysis of the durational properties of this selection of 35 triplets was conducted [7]. A re-analysis of these data with a single repeated measures factor across sequences shows a main effect of the boundary condition on C and V1 durations. Duration of pre- and post-consonantal vowels.

Compared to the syllable-boundary condition (cond. C), the duration of the pre-consonantal vowel (V1) is significantly longer in enchainement condition (cond. A; F(1,279)=57.6, p<.0001) and in word-boundary condition (cond. B; F(1,279)=44.41, p<.0001). By contrast, the comparison of V2 duration does not show any significant effect of the boundary condition.

Duration of consonant:

The duration of the consonant also varies according to the boundary condition. Indeed, the consonant is significantly shorter in the enchainement condition (cond. A) than in both word-initial (cond. B; F(1,279)=38.87, p<.0001) and syllable-onset (cond. C; F(1,279)=13.96, p<.001) conditions. Moreover, the word-initial consonants were significantly longer than the syllable-onset consonant (F(1,279)=5.9, p<.05).

In sum, the pre-consonantal vowel was significantly longer when it constitutes the nucleus of a word-final syllable (A: V1C# or B:V1#) compared to when it is word-initial (C: V1.CV2). Furthermore, the consonant was significantly shorter when it constitutes the coda of a word-final syllable (in A compared to B and C) or the onset of the initial syllable (B vs. C). The durational properties of the enchainement consonant differed from those of syllable onsets, or word-initial consonants (5), [6], [7], [9], [11]). Thus, these results confirm the presence of subtle acoustic differences between the three boundary conditions. Overall speakers, these differences are found on “V1 + C” sequences.

Figure 1: Average duration for 8 speakers of pre-consonantal vowel (V1), inter-vocalic consonant (C) and post-consonantal vowel (V2) in three boundary conditions.

Nevertheless, as shown in Fougeron et al. [7], this effect of the boundary condition depends upon the speaker. Indeed, boundary condition effects on C and V1 duration are not present for all the speakers (at most 5 of the 8 speakers), sometimes present on the pre-consonantal vowel (V1) but not on the enchainement consonant (C) or vice-versa (see [6] for details). Hence, for the perception studies presented in the next section, we selected a speaker who matches the overall data well (Figure 2). Indeed, for this speaker, the pre-consonantal vowel (V1) was significantly lengthened in conditions A and B, i.e. when it constitutes the nucleus of a word-final syllable (respectively, A vs. C: F(1,34)=27.6, p<.0001; B vs. C: F(1,34)=17.9, p<.001). Concerning the consonantal duration, this speaker shows a significantly shorter enchainement consonant (cond. A) compared to word-initial consonant (cond. B; F(1,34)=10.6, p<.01) and a shorter enchainement consonants compared to syllable-onset consonants (cond. C; F(1,34)=4, p=.05). Thus, this speaker differentiates between the three boundary conditions on either V1 or C duration or both.
3. Perception data

In order to test the perceptual relevance of the phonetic cues found in the production study, two experiments were run. We used a same-different task in which participants had to decide as quickly and as accurately as possible whether the members of the pairs were the same or different. Since we were interested in the participants’ ability to discriminate between the three types of sequences, we used percent correct data rather than RTs.

3.1. Experiment 1

3.1.1. Method

Participants: Twenty students at the University of Geneva participated in the experiment for course credits. All were native speakers of French and reported no hearing disorders.

Materials & Procedure: VCV sequences were extracted from 33 of the triplets and presented pair wise to the participants. Participants received 198 pairs. Half of the pairs consisted of two identical sequences and the other half consisted of two sequences extracted from different boundary contexts. The “same” trials were constructed by presenting the same productions extracted from two different boundary conditions (i.e. A vs. A, B vs. B, C vs. C). The “different” trials consisted in two different sequences extracted from two different boundary conditions (i.e. A vs. B, B vs. C, A vs. C).

A beep was first played to indicate the beginning of the trial. Participants were then presented with two sequences separated by an interstimulus interval (ISI) of 400 ms. They were instructed to make their decision as quickly and as accurately as possible. If participants failed to respond within 4000 ms of the onset of the stimulus, no response was recorded and the next item was presented. An intertrial interval of 1000 ms elapsed between the end of one trial and the beginning of the next. Half of the participants responded “same” with their right hand, and the other half with their left hand. They began the experiment with 12 practice trials.

3.1.2. Results and discussion

The results are presented in Table 2. For the “same” responses (AA, BB and CC conditions), performance was high and above chance according to the t-tests. Of greater interest is participants’ performance on the different trials (AB, AC and BC conditions). Performance in each of the three experimental conditions was above chance both, in analyses by participants and by items (AB: t1(19) = 2.99; p<.01; t2(32) = 1.88; p<.05; AC: t1 (19) = 5.22; p<.001; t2(32) = 4.16; p<.001; BC: t1 (19) = 5.92; p<.001; t2(32) = 3.44, p<.001), thus indicating that participants were able to discriminate between the three types of sequences.

<table>
<thead>
<tr>
<th>SAME</th>
<th>DIFFERENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>BB</td>
</tr>
<tr>
<td>% Correct</td>
<td>Responses</td>
</tr>
<tr>
<td>94</td>
<td>92</td>
</tr>
</tbody>
</table>

In sum, Experiment 1 suggests that French listeners are sensitive to subtle acoustic cues due to enchaînement. Indeed, listeners are sensitive to both the nature - syllabic or lexical (cond. A vs. C) - and the position - before/after the consonant (cond. A vs. B) - of the boundary. Hence, these results support the view that subtle phonetic cues to different types of boundaries can be used in a syllable-based segmentation heuristic. However, it is not clear whether the “different” responses were made on the basis of physical similarity or more high-level differences between the two target sequences. Experiment 2 aims at testing whether discrimination resists low-level acoustic variations. Unlike experiment 1, the identical sequences were physically different and were two different productions by the same speaker.

3.2. Experiment 2

3.2.1. Method

Participants: Twenty additional students at the University of Geneva participated in the experiment for course credits. All were native speakers of French and reported no hearing disorders.

Materials & Procedure: They are the same as in Experiment 1 except that two different productions by the same speaker were used for the identical sequences. Thus, the “same” responses were made on sequences extracted from two different productions of the same speaker. In other words, the “same” trials consisted in two different productions of the same boundary condition (i.e. A vs. A, B vs. B, C vs. C). The “different” trials again consisted in two different productions extracted from two different boundary conditions (i.e. A vs. B, B vs. C, A vs. C).

3.2.2. Results and discussion

<table>
<thead>
<tr>
<th>SAME</th>
<th>DIFFERENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>BB</td>
</tr>
<tr>
<td>% Correct</td>
<td>Responses</td>
</tr>
<tr>
<td>77</td>
<td>76</td>
</tr>
</tbody>
</table>

Results are presented in Table 3. For the “same” responses (AA, BB and CC conditions), performance decreased as compared to Experiment 1 but was still above chance. For the “different” responses (AB, AC and BC conditions), performance also decreased and was above chance in the BC condition.
condition only in the by participants analysis ($t_1(19) = 2.37; \ p<.05; \ t_2(32) = 1.04, \ p=.15$). Performance was not above chance in the \textit{AB} ($t_1(19) = -1.41; \ p=.09; \ t_2(32) = -.71, \ p>.20$) or \textit{AC} ($t_1 (19) = 1.35; \ p=.10; \ t_2(32)= 1.07, \ p=.15$) conditions.

4. Conclusions

The aim of this paper was twofold. First, we tested whether durational cues that distinguish sequences sharing the same segmental content but differing with respect to the position (before/after C) and the type (syllabic/lexical) of boundaries can be measured. Second, we tested whether native French listeners were able to use these acoustic cues to differentiate between different types of boundaries.

Our results provide further evidence that so-called resyllabification in enchaînement is not always complete in French. Thus, it appears that enchaînement in French does not neutralize the contrast between a sequence which contains a word-final coda consonant (V1.C#V2) and a sequence in which the same critical consonant is the onset of the following syllable (V1#C#V2). Compared to the syllable-boundary condition, the enchaînement condition was found to differ in both (1) a longer pre-consonantal vowel (V1) and (2) a shorter consonant (C). Moreover, when compared to the word-boundary condition, the enchaînement condition differs in a lengthened pre-consonantal vowel and a shorter consonant. Therefore, it appears that the enchaînement consonant does not adopt the durational properties of either syllabic onsets, or word-initial consonants. These results also challenge models of speech production assuming resyllabification at the level of phonological syllables (e.g. [12]). Indeed, the production of V1#C V sequences depends upon articulatory gestures different from those required for the production of their V1#CV or V.C counter-parts.

The perceptual data observed in the same - different task are somewhat disappointing, since they suggest that listeners do not always use fine-grained phonetic cues to differentiate between the three phonetic sequences tested (A, B and C). In Experiment 2, where low-level acoustic variations were introduced, participants were able to use acoustic cues only to distinguish between word and syllable boundaries. Nonetheless, since the difference is only significant in the by participant analysis, it must be interpreted with caution and confirmed by further research. No discrimination was made between enchaînement and the other two boundary conditions. Note however that although the available acoustic cues are not salient enough to differentiate two different boundary conditions, they allow participants to regroup two different productions of a same boundary condition.

The results found on the “different” trials do not parallel the data from word spotting experiments [5] in which listeners appear to use durational differences to distinguish between different syllabifications. However, the discrepancies between our results and other findings available in the literature could be related to the linguistic stimuli. Indeed, we used very short stimuli (average duration 350 ms), which probably make up an insufficient context to perform the task accurately. Longer sequences might be required for listeners to be sensitive to the durational information. Additionally, speech samples used for perceptual experiments were extracted from one talker only, which makes it difficult to generalize our findings. Thus, more research looking at the effect of the stimuli duration is required to solve the apparent divergence with past studies. Moreover, it could be especially interesting to use speech samples from a speaker who exhibits spectral differences on vowels V1 and/or V2 to test whether listeners use spectral cues to differentiate between the three phonetic sequences tested.

To sum up, this study confirms that enchaînement does not neutralize the contrast between the different types of boundaries studied. Moreover, the preliminary perception experiments indicate that listeners are able, although to a limited extent, to use durational cues. Further experiments, and particularly comparisons of the effects shown by several tasks involving different processes such as word-spotting [5] are required to assess how and when French listeners use fine-grained phonetic cues to avoid missegmentations.

5. Acknowledgements

This research was supported by grant n° 1114-059532/5 from the Fond National de Recherche Suisse (FNRS).

6. References


