ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

PHONOLOGICAL CONSTRAINTS IN SPEECH SEGMENTATION PROCESSES: INVESTIGATING LEVELS OF IMPLEMENTATION.

Crouzet, Olivier & Bacri, Nicole

Laboratoire de Psychologie Expérimentale
UMR 8581 - CNRS / Université René Descartes
28, rue Serpente - 75006 Paris - France
E-mail: crouzet@psycho.univ-paris5.fr

10.21437/Eurospeech.1999-673

ABSTRACT

Recent data have been considered as evidence for a role of phonological constraints in speech segmentation processes. A distributional analysis of consonant sequences in the French lexicon shows that these results may be accounted for by lexical competition or by transitional probability models in which no claim is made about the usefulness of phonology in speech perception. Two word-spotting experiments were devised in order to investigate this issue. In the first experiment, only half of the participants showed the previous effect. In a modified replication of this study, the expected effect emerged clearly with the same target stimuli. While the role of phonotactics in speech segmentation seems to be independent from lexical competition and transitional probabilities, the likely role of strategical effects is discussed.

INTRODUCTION

It has been argued that knowledge about the phonological constraints that govern a language may be used in speech segmentation. Recent data ([7], [10]) obtained with various experimental paradigms have been interpreted as evidence for such processes. In these experiments, so called legal phoneme sequences were contrasted with illegal ones in experiments involving the word-spotting [5] or the phoneme monitoring task.

McQueen [7] used a word-spotting task in which Dutch speakers had to spot real words embedded in disyllabic non-words (e.g. detecting *pil*, pill, in [pɪlvrem]). In one condition, words occurred in initial position with the two following consonants giving rise to either a legal (e.g. /vr/ in [pɪlvrem], with the target word *pil*, pill) or an illegal consonant cluster (e.g. */mr/ in [pɪlmrem]). When the consonant cluster was phonotactically legal, the syllabic and lexical boundaries were aligned. An illegal consonant cluster gave rise to a misalignment of these boundaries. In a second condition, words occurred in final position with the *one* preceding consonant giving rise to either a legal (e.g. /dr/ in [fidrɒk], with the target word *rok*, skirt) or an illegal consonant cluster (e.g. */mr/ in [fimrɒk]). In this condition, the relationship between consonant clusters legality and the alignment of syllabic and lexical boundaries was reversed; boundaries being aligned for illegal clusters and misaligned for legal ones. Listeners were both more accurate and faster at detecting words when they were *aligned* with the syllabic boundary than when the lexical and syllabic boundaries were *misaligned*. This result was interpreted as an evidence of the influence of phonotactic constraints upon lexical segmentation processes.

Vroomen & De Gelder [10] used a phoneme monitoring task in which Dutch speakers listened to sentences like 'de boot die gezonken is...' or 'de boot is gezonken...'. The participants' task consisted in detecting a prespecified phoneme. In the examples given above, the target phoneme /t/ was either in syllable-coda position ('de.boo*t*.die.ge.zon.ken', with the target obstruent followed by another obstruent) or in syllable-onset position ('de.boo.*t*is.ge.zon.ken', with the target phoneme followed by a vowel). When the target phoneme occurred as a coda, the lexical and syllabic boundaries were aligned; whereas they were misaligned when it was pronounced as the onset of a syllable. As was observed in McQueen's [7] word-spotting experiments, the phoneme monitoring latencies were slower when the boundaries were misaligned (target phoneme in onset) than when they were aligned (target phoneme in coda).

PHONOLOGICAL WELL-FORMEDNESS VS. LEXICAL FREQUENCY

As a matter of fact, legal and illegal clusters, as well as obstruent-vowel and obstruent-obstruent sequences, lead to contrasting syllabic structures. The effects observed in these experiments may consequently reflect the requirement of phonological constraints in speech segmentation processes. However, legal and illegal consonant sequences are also likely to differ with respect to their frequency of occurrence in the target language. In the French language, a distributional analysis of consonant clusters was performed on a 35-thousand-words phonemic computerised database (BRULEX). Frequency of occurrence of consonant clusters was weighted with lexical frequency and translated to a logarithmic scale. This analysis showed that, without taking the location of phoneme sequences in the words into account, legal clusters occur much more frequently than illegal ones. For example, /vr/ which is also told to be legal in French has a logarithmic frequency of 77 whereas */mr/ only exhibits a frequency of 43. It is worth noting that the effects observed by [7] and [10] were also obtained with French speakers listening to French materials ([1], [4]) and that, in our analysis, all the legal clusters spanned the high end of the scale whereas most of the illegal ones spanned the low end of the scale. We did not have the opportunity to perform the same analysis on a Dutch lexicon but we argue that a very similar pattern may occur. If this were actually the case, any model in which frequency could be implemented may predict these data.

The first kind of model that could account for these effects is grounded on transitional probabilities ([2], [3], [9]) and would predict that high probability strings are likely to belong to the same word whereas low probability strings would rather span lexical boundaries. Of course, one may argue that this kind of procedure would be close to a speech segmentation process founded on phonological knowledge as this kind of behaviour would be implemented at a prelexical stage of processing and that this stage would lie on distributional regularity. Nevertheless, the analysis performed on the French language showed that some so-called illegal consonant clusters (e.g. */ks/), while hardly pronounced at the onset of a word, are rather frequent inside and at the end of words (and sometimes as frequent as legal clusters, e.g. /kr/). Therefore, we would not expect the same effects depending on the clusters we take into account and the interpretation we favour.

Lexical competition processes may also involve the *implicit* use of phonemic strings' frequency at a lexical stage of processing. Indeed, when a frequent string occurs in the speech signal, the behaviour of a lexical competition model consists in triggering the activation of many words in its lexicon whereas a rare phonemic string would lead to the activation of only few words. In [7] and [10], we may argue that due to a higher number of activated competitors when a legal sequence occurs, it would be harder and longer to identify a word in the aligned condition than in the misaligned one. The effects that were interpreted as evidence for a role of phonological constraints in speech segmentation processes may consequently be accounted for by a lexical competition model like Trace [6] or Shortlist [8] without the need to involve any kind of phonological knowledge. Again, one may argue that in [7], when the to-be-detected words occurred at the onset of the sequence, the *legal* (i.e. frequent) cluster gave rise to an aligned condition. Therefore, a lexical competition account would have led to predict slow reaction times for the legal-aligned condition whereas in this condition, reaction times were actually faster than in the illegal-misaligned condition. However, a control experiment showed that some acoustical differences between words in the aligned and the misaligned conditions may have accounted for the original data; we may consequently not trust this part of the results to interpret the effects.

It seems that there may be interpretations of these effects other than those proposed by McQueen [7] and Vroomen & De Gelder [10]. These results may be explained by any model of speech segmentation and stronger controls are needed to allow for a decisive account of these so-called phonotactical or syllabification effects. The two experiments described in this paper aimed at testing between these interpretations.

EXPERIMENT 1

The aim of this experiment was to dissociate the role of phonological constraints and either lexical competitions or transitional probabilities in the emergence of the effects observed in [7] and [10].

Method

Participants

Forty French speaking students took part in this experiment for course credit. None had experienced any auditory impairment.

Materials

From the distributional analysis that we performed on the French lexicon, we selected three pairs of legal and illegal consonant clusters that exhibited the same frequency. Twenty monosyllabic words were pronounced at the beginning of disyllabic nonsense sequences and digitised at 16 bit with a sampling frequency of 16 kHz. The final consonant of the target words was always a stop. The following consonant gave rise to either a legal (e.g. /gl/) or an illegal (e.g. /gz/) cluster, both being equally frequent. Twenty filler stimuli (e.g.: /tralbɔd/) were also recorded in which no word occurred. In the legal condition, lexical and phonological boundaries were misaligned; they were aligned in the illegal condition. Speech strings in legal and illegal conditions only differed with respect to the second consonant of the medial cluster. As legal and illegal clusters had the same frequency of occurrence in the lexicon, the same number of lexical competitors should be activated by the speech perception system. With this kind of manipulation, we may predict that only if the effects observed in [7] and [10] were actually due to the use of phonological constraints should they replicate in a similar word-spotting task ([5], [7]). Conversely, if the effects have to be explained in terms of transitional probabilities or lexical competition processes, they should not replicate here.

Table 1: Examples of stimuli used in Exp. 1 and 2.

legal	illegal
bague lure	bague zure
/baglyr/	/bagzyr/
mean log-freq. = 81	mean log-freq. = 81.7

Procedure

We used the word-spotting task. French speaking listeners were asked to press a button as soon as they would detect monosyllabic words occurring at the beginning of nonsense strings. Target words always occurred at the beginning of the strings. Each participant heard stimuli in the legal and the illegal condition without listening to the same word twice. Participants were asked to say the word they had heard aloud after pressing the button. Reaction times were measured from the burst of the word's final consonant. The experiment began with a familiarisation phase.

Results

Participants that detected more than 50% of the words (30 subjects) were selected for the analysis of variance. Reaction times greater than 1720 ms were rejected. No significant effect emerged from the data. Reaction times were similar in the legal (835 ms) and the illegal (865 ms) conditions (All F's were non significant in the analyses by subjects and by items). This seems to confirm the role of lexical competitions or transitional probabilities in the effects observed in [7] and [10]. However, a closer look at the data showed that some participants exhibited the previous effect while the effect was reversed for some others. This was not due to a 'list' effect as the 'list' variable did not interact with the 'legality' variable. Distinct analyses were conducted for 13 subjects on one hand (with the expected effect) and 17 subjects on the other hand (with the reversed effect). These data are displayed in Figure 1. In the first 'group' of subjects (for which the expected effect emerged), the legality effect proved to be significant by subject (legal: 762 ms; illegal: 825 ms; $F_1(1,11) = 6.845$, $p < .05$) but also by item ($F_2(1,19) = 8.152$, $p = .01$), confirming that this effect was not a 'list' effect. In the second group (for which the effect was reversed), subjects were faster to detect the words in the legal condition (i.e. misaligned, 834 ms) than in the illegal one (i.e. aligned, 945 ms). This effect was significant by subject ($F_1(1,15) = 15.117$, $p = .001$) but was only marginally significant by item ($F_2(1,19) = 3.507$, $p = .077$). Moreover, in the analysis by subjects, the 'legality' and 'list' variables interacted.

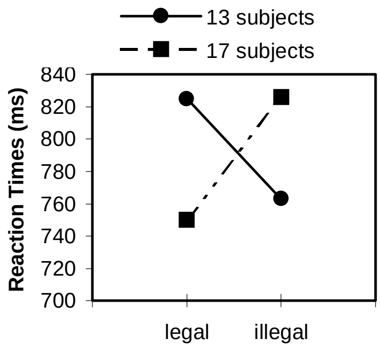

Figure 1: Two directions of the effect depending on the subjects taken into account (Reaction times in ms).

Discussion

It seems that the effect observed by McQueen [7] and by Vroomen & De Gelder [10] may replicate despite a better control of the frequency of consonant clusters and lexical competitions. However, this effect emerged for only about one half of the listeners (13 participants among 30). If we look at the results for the 30 subjects, no effect emerged at all. Nevertheless, for those exhibiting the effect predicted by a phonological constraints account, it appeared to be rather reliable. One possibility would be that the restrictions on the consonant clusters that were compared in the experiment may have led us to select particular target words that would have made the task harder than in previous studies. Due to the intrinsic difficulty of the word-spotting task and to the specific words used in this experiment, subjects may have tried to develop strategies allowing them to achieve the best possible performance. One strategy may consist in segmenting the phonemic string into syllables. This strategy would lead to the emergence of the previous effects. Another way of performing the task could consist in achieving a pure phonemic analysis and to try to identify a word from this analysis, without any syllabic processing. While listening to the stimuli, it seemed to us that both kinds of obstruents were rather different. Obstruents in an obstruent-liquid cluster may be more ‘prototypical’ than those pronounced in an obstruent-fricative one. It may then become harder to identify this obstruent in one context than in the other, which could explain the inversion of the effect for the 17 remaining participants. We then decided to replicate the first experiment with an alternate procedure in order to minimise its difficulty.

EXPERIMENT 2

The second experiment was the same as the first one, except that legal and illegal clusters were not mixed in the experimental lists. This was intended to make the experiment easier to perform as participants would not randomly turn from one structure (e.g. illegal-aligned) to the other (e.g. legal-misaligned).

Method

Participants

Twenty-eight subjects ran this experiment. They were similar to those of experiment 1.

Materials and procedure

The same task and stimuli were used in both experiments. Only the procedure differed. In the 2^{nd} experiment, subjects first listened to stimuli in one condition (e.g.: legal) for half of the materials, and then listened to stimuli in the other condition for the other half. The order of presentation of legal and illegal stimuli was counterbalanced across subjects. As in experiment 1, no subject had to detect the same word twice.

Results and discussion

Twenty participants detected more than 50% of the words and were selected for the analysis of variance. Three words were missed by more than 50% of the participants and were rejected from the analysis. Reaction times greater than 1740 ms were not taken into account. No interaction occurred between the ‘legality’ variable and respectively the ‘list’ and ‘order’ variables. Words were detected faster in the legal condition (871 ms) than in the illegal one (811 ms). This effect was significant by subjects ($F_1(1,19) = 4.811$, $p < .05$) but only marginally significant by items ($F_2(1,16) = 3.537$, $p < .078$). A slight modification of the experimental procedure led to rather reliable effects with contrast to what emerged from the data of experiment 1.

GENERAL DISCUSSION

These data seem to confirm the role of phonological constraints in speech segmentation processes as was argued by [7] and [10]. However, depending on the procedure used in the experiment, it was not always clear whether the effect was reliable. In the first experiment, it is worth noting that, contrary to what was observed in previous studies and in the second experiment, more than half of the participants gave slower responses in the illegal than in the legal condition. If a slight modification of the experimental procedure can change the data dramatically, one may wonder whether these so-called phonological effects would not be better described in terms of strategical effects.

BIBLIOGRAPHICAL REFERENCES

[1] Bacri, N., & Banel, M.-H. (1995). Do metrical and phonotactic cues cooperate in spoken word recognition? *XIIIth ICPhS*, *3*, 608-611.

[2] Brent, M.R. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research*, *26*, 363-375.

[3] Brent, M.R., & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93-125.

[4] Crouzet, O. & Bacri, N. (1998). Groupes consonantiques illégaux et segmentation de la parole. *XXII^nd Journées d’Etudes sur la Parole*, 275-278, Martigny, Switzerland.

[5] Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *JEP: HPP*, *14*, 113-121.

[6] McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

[7] McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *JML*, *39*, 21-46.

[8] Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189-234.

[9] Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *JML*, *35*, 606-621.

[10] Vroomen, J., & De Gelder, B. (forthcoming). The role of resyllabification in lexical access. *Memory & Cognition*.