Proceedings of the Institute of Acoustics

ON THE IMPLEMENTATION OF PHONOLOGICAL CONSTRAINTS IN COMPUTATIONAL MODELS OF SPEECH IDENTIFICATION

O. Crouzet MacKay Institute of Communication and Neuroscience, Keele University.
N. Bacri Laboratory of Experimental Psychology, Paris 5 University, France.

1 INTRODUCTION

It has been argued that phonological regularities may be used by native speakers to locate word boundaries in the speech signal [2]. Evidence of the influence of phonotactic or syllabic regularities on speech identification processes has been found with Dutch native-speakers [8], [6]. As French phonology bears similar syllabification processes, these kinds of regularities may be useful in French as well. However, a distributional analysis of consonant clusters in the French lexicon shows that there is a strong relationship between phonological regularity and frequency of occurrence in the lexicon. Various cognitive processes may therefore account for the observed data. A series of experiments was conducted in order to test between these interpretations. This issue was investigated with two tasks (word-spotting and phoneme detection) which show discrepancies concerning the possible influence of either phonology or frequency in word identification processes. As these experimental paradigms may involve alternative processes, this difference may be analysed in terms of the various influences that phonological constraints and probabilistic regularities may have on speech recognition. This observation enables a discussion to be developed concerning the relationship between prelexical segmentation processes and word activation level.

1.1 Previous work

The influence of phonological regularities on speech identification processes has been evidenced using two different tasks (word-spotting [6] and phoneme detection [8]). In each of these studies, an alternative aspect of sequential phonological regularities was investigated.

In the *word-spotting* task [6], participants had to detect real monosyllabic words embedded at the beginning or at the end of nonsense bisyllabic sequences (e.g. *pil* –pill– in /pilvrem/). In the initial condition, the two consonants which followed the word gave rise to either a phonotactically legal (e.g. /vr/ in /pilvrem/) or illegal (e.g. */mr/ in /pilmrem/) consonant cluster. When the consonant cluster was phonotactically legal, syllabic and lexical boundaries were aligned. An illegal consonant cluster gave rise to a misalignment of these boundaries. When targets occurred in *final* position, it was the relationship between the word's initial consonant and the preceding phoneme which accounted for the alignment variable (e.g. /dr/ in /fidrok/, with the target word *rok* –skirt– vs. */mr/ in /fimrok/). The relationship between consonant clusters' legality and phonotactic–lexical alignment was then reversed. Error rates were higher when a misalignment occurred between phonotactic and lexical boundaries.

Vroomen & De Gelder [8] investigated a similar issue with a phoneme detection task in which speakers had to monitor target phonemes inside sentences. The target was either pronounced at the coda (in 'de.boo*t*.die.ge.zon.ken [...]') or at the onset (in 'de.boo.*t*is.ge.zon.ken [...]' of the syllable which followed the monosyllabic word. In the onset condition, there was a misalignment between the syllabic and lexical boundaries; these were aligned in the coda condition. Longer reaction times were observed when the target phoneme was in onset position than when it was pronounced in coda. Rather than referring to phonotactic constraints, Vroomen & De Gelder use the concept of syllable structure to account for their data.

Though the respective authors of these two studies refer to different concepts (phonotactics and syllabification), their data may be accounted for by a similar process: sequential phonological constraints.

1.2 Phonotactics and syllabification

Languages are structured by a set of constraints which any speaker has integrated in the course of language acquisition. Phonotactic regularities are often defined as constraining the phoneme sequences which may be pronounced at the beginning of a word. For example, /tl/ is considered to be phonotactically illegal in many western languages; it is prohibited at the beginning of words but may perfectly legally occur inside words (e.g. *atlantic*). These may be viewed as *linear* regularities constraining the organisation of sound sequences. Alternatively, syllabic structure is usually considered as a *hierarchical* organisation of phonemes within the spoken stream. However, syllable organisation is mostly governed by phonotactic constraints. The previous data may therefore be accounted for by a common process of phonological analysis of the input stream leading listeners to segment the speech signal on the basis of their knowledge about sequential phonological organisation.

1.3 Phonology and frequency

A distributional analysis of the occurrence of consonant clusters in the French lexicon has been performed on a 35,000 words computerized dictionary [1]. Stop-initial consonant clusters' frequency was analysed and categorized on the basis of the phonetic characteristics of the following consonant (liquid, stop, fricative). It was found that tautosyllabic–phonotactically legal sequences (stop–liquid, e.g. /tr/) tend to be more frequent than heterosyllabic–phonotactically illegal (stop–fricative, e.g. /ts/ or stop–stop, e.g. /tk/) sequences when their frequency is computed wherever they occur in the language's words. It is therefore not clear whether the previous data may be accounted for in terms of phonological organisation of the input stream or, alternately, on the basis of probabilistic computation. Though there is a strong relationship between phonological structure and frequency, the observed distributions overlap. It was therefore possible to dissociate the respective contribution of phonological organisation and probabilistic computation with a specifically designed set of consonant sequences. A series of 4 experiments was conducted in order to investigate this issue.

2 WORD SPOTTING TASK

The respective influence of phonology and frequency was first investigated with a word-spotting [7] task in which participants have to detect real words embedded in bisyllabic nonsense strings.

2.1 Experiment 1: Phonological influence

2.1.1 Participants

Twenty-four French speaking students took part in the experiment for course credit. None had experienced any auditory impairment.

2.1.2 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. Eighteen monosyllabic words were pronounced at the beginning of bisyllabic nonsense sequences and digitized on 16 bits at 16 kHz. The final consonant of the target was always a stop. Target words always occurred at the beginning of the nonsense sequences. The following consonant gave rise to either a tautosyllabic (e.g. /gl/) or a heterosyllabic (e.g. /gz/) cluster of similar frequency. Twenty filler stimuli (e.g. /tralbyd/) were also recorded which did not contain an actual word of the French lexicon. In the tautosyllabic condition, lexical and syllabic boundaries were misaligned (/va#gryn/, target /vag/, *vague*, wave); they were aligned in the heterosyllabic condition (/vag#zyn/).

2.1.3 Procedure

Listeners were asked to press a button as soon as they could detect an unspecified word embedded at the beginning of the nonsense stimuli. Participants had to pronounce the word they identified after each positive answer. Reaction times were measured from the burst of the first consonant in the consonant cluster (final consonant of the target word).

2.1.4 Results and discussion

Participants who detected more than 50% of the words (20 participants) were selected for the analysis of variance. Reaction times greater than the mean plus 2 standard-deviations (i.e. RT greater than 1720ms) were not taken into account.

Higher reaction times were observed when the final consonant was pronounced at the onset of the following syllable than when it was in coda position. This effect is statistically significant in the subjects analysis ($F_1(1,19) = 6.022$, $p < .05$) but only marginally significant in the items analysis ($F_2(1,16) = 4.216$, $p = .057$). However, both analyses provide statistically significant differences when computed over the half most frequent words of the sample ($F_1(1,19) = 8.134$, $p < .05$; $F_2(1,8) = 6.130$, $p < .05$). The corresponding measurements are displayed in Table 1. It is therefore possible to find a

Table 1: Reaction times (standard-error) and error rates observed in Experiment 1. Word-spotting task, frequency-matched consonant clusters. Data computed over the half most frequent words in the sample.

	tautosyllabic	heterosyllabic
reaction times (ms)	833 (45)	746 (43)
error rates (%)	15.5	13.0

phonological effect without any frequency difference between the consonant clusters involved in the comparison. However, as the effect only clearly occurred for the most frequent words, the contribution of frequency alone needs to be investigated.

2.2 Experiment 2: Probabilistic Influence

2.2.1 Participants

Participants were the same as those who took part to Experiment 1.

2.2.2 Materials

Alternate pairs of consonant clusters were selected which had a similar phonetic structure (stop–fricative or stop–stop) but differed in terms of their frequency of occurrence in the French lexicon.

2.2.3 Procedure

The procedure was similar to that used in the first experiment. Order of presentation of Experiments 1 and 2 was counterbalanced between subjects.

2.2.4 Results and discussion

The same criteria were applied to select participants on the basis of their performance. Reaction times greater than 1600 ms were not taken into account.

There is no evidence of a frequency effect in the reaction times analysis ($F_1 < 1$; $F_2 < 1$). Nevertheless, frequent consonant clusters give rise to higher error rates than rare ones. This effect is significant in the subjects analysis ($F_1(1,19) = 4.930$, $p < .05$) but only marginally significant in the items analysis ($F_2(1,26) = 3.250$, $p = .083$). The observed data are depicted in Table 2. Though a

Table 2: Reaction times (standard-error) and error rates observed in Experiment 2. Word-spotting task, phonologically similar consonant clusters.

	frequent	rare
reaction times in ms (SE)	767 (36)	758 (36)
error rates (%)	22.0	16.8

slight effect of frequency may occur in the error rates analysis, nothing seems to occur in terms of speed of processing and this effect is only clearly significant in the subjects analysis.

2.3 General Discussion

It is quite difficult to make sense of these data as each of these experiments gave rise to marginally significant comparisons in the items analysis. Reducing the analysis to a subset of the materials enables this comparison to reach the significance level in Experiment 1. Therefore, phonological segmentation of the speech input may involve actual phonological processing rather than probabilistic computation. Indeed, comparing tautosyllabic and heterosyllabic consonant clusters leads to longer reaction times when phonological and lexical boundaries are misaligned than when they are aligned. This holds even when frequency of occurrence is kept constant.

Yet, the effect does not reach significance when the whole set of words is taken into account. As a matter of fact, the word-spotting task proves to be quite difficult [2], [7]. This difficulty may have induced participants to focus processing on some class of information which may have been easier to use than phonological regularities in order to identify words in the speech input. As this phonological effect had already been observed in the easier phoneme detection task [8], a similar investigation was conducted with this alternative task.

3 PHONEME DETECTION TASK

The respective contribution of phonology and frequency in speech identification processes was investigated with another experimental paradigm in order to reduce the difficulty of the experiment.

3.1 Experiment 3: Phonological influence

3.1.1 Participants

Twenty-four French speaking students took part in this experiment. None of them had participated in experiments 1 and 2.

3.1.2 Materials

Using the phoneme detection task involved a modification of the stimuli in order to (1)prevent the occurrence of a target phoneme more than once and (2)use stimuli which would be closer to what one is used to process in everyday communication. Short *article + noun + adjective* sequences were recorded. The combination of the noun's final consonant and of the adjective's initial one gave rise to either a tautosyllabic (e.g. /ȳsa*kl*avabl/, *a washable bag*) or a heterosyllabic (/ȳsa*kt*ise/, *a sewed bag*) cluster. As in Experiment 1, tautosyllabic consonant clusters induced a misalignment between phonological and lexical boundaries.

3.1.3 Procedure

Listeners were asked to press a button as soon as they could detect a prespecified phoneme in the stimuli. As in the word spotting task, reaction times were measured from the burst of the first consonant in the relevant consonant cluster (final consonant of the noun).

3.1.4 Results and discussion

Contrary to what was found in Experiment 1, neither reaction times nor error rates analysis gave rise to an effect of phonological structure when frequency was kept constant ($F_1 < 1$; $F_2 < 1$). The corresponding data are displayed in Table 3. In order to complete the task comparison, a complementary investigation was conducted in order to study the contribution of frequency alone in the phoneme detection task.

Table 3: Reaction times (standard-error) and error rates observed in Experiment 3. Phoneme detection task, frequency-matched consonant clusters.

	tautosyllabic	heterosyllabic
reaction times (ms)	604 (41)	598 (34)
error rates (%)	3.1	7.3

3.2 Experiment 4: Probabilistic influence

3.2.1 Participants

Participants were the same as those who took part to Experiment 3.

3.2.2 Materials

Similar *article + noun + adjective* sequences were selected with alternate pairs of consonant clusters. These had a similar phonetic structure (stop–fricative or stop–stop) but differed in terms of their frequency of occurrence in the French lexicon.

3.2.3 Procedure

The procedure was the same as in Experiment 3. Order of presentation of Experiments 3 and 4 was counterbalanced between subjects.

3.2.4 Results

Contrary to what was observed in the word-spotting task, a clear effect of frequency was observed, frequent consonant clusters giving rise to slower reaction times than rare consonant clusters ($F_1(1,23) = 6.371$, $p < .05$; $F_2(1,11) = 10.716$, $p < .01$). Data are displayed in Table 4.

3.3 General Discussion

As a mirror image of the data pattern observed in the word-spotting task, phonological regularities do not seem to influence processing in the phoneme detection task whereas frequency of occurrence

Table 4: Reaction times (standard-error) and error rates observed in Experiment 4. Phoneme detection task, phonologically similar consonant clusters.

	frequent	rare
reaction times (ms)	501 (25)	465 (21)
error rates (%)	4.3	6.1

does. In the word-spotting task, phonological organisation only has an effect on a subset of the lexical material, this effect can not however be accounted for by a probabilistic interpretation which would involve a lexical level of processing as consonant clusters' frequency is kept constant. In the phoneme detection task, the observed effects are easier to analyse as they stand for the whole set of stimuli. However, basing interpretation only on this set of experiments would lead to confirmation of the hypothesis that only probabilistic computation is performed during speech identification, with no real contribution of phonological regularities.

4 DISCUSSION

Though the task comparison which has been performed here may seem to provide contradictory results concerning the influence of phonological regularities vs. frequency on speech identification processes, these discrepancies may help to constrain an analysis of the relationship between the various levels of processing involved in the identification of linguistic signals. We first discuss the reasons why phonology and frequency may influence speech recognition differently with respect to the experimental task. We then turn to an analysis of the relationship between prelexical and lexical levels of processing, and make use of the preceding discrepancies to illustrate this issue.

4.1 Task discrepancies

There are several reasons why phonology and frequency may influence processing differently depending on the experimental task.

1. The word-spotting task may be too difficult to enable participants to process the speech input in a natural way. Therefore, they may have used artificial processes which would not be involved in everyday communication. As a matter of fact, segmenting the input in syllabic-sized units may have helped locating word boundaries when phonological and lexical boundaries are aligned. These segmentation procedures may well be absent when processing *natural speech*.

2. The phoneme detection task may not be appropriate for investigating the influence of phonological constraints on the location of word boundaries. Indeed, though it has been shown that

lexical information may be useful in this task [3], detecting a phoneme does not necessarily imply the identification of the word in which it appears. A bottom-up acoustic-phonetic analysis may be sufficient to locate a phoneme in the speech signal. As a matter of fact, if phonological regularities were to influence the processes involved in the phoneme detection task, they would necessarily do so by having an effect on word identification. Only by doing so would the data observed be dependent upon phonological organisation.

3. There is however a third interpretation which is related to the way prelexical and lexical information are integrated during speech processing. Both tasks may make it possible to investigate natural cognitive processes, the observed discrepancies being predicted by a model in which lexical information is independent from prelexical stages of phonological segmentation. This interpretation is depicted in the following section.

4.2 Information integration in speech recognition

In every field of cognitive science, low–level processes are usually considered as available to influence higher–level processes. The study of word identification processes has led to similar arguments in attempts to model the role of phonological constraints in the segmentation of speech into words. Usually, prelexical processes of phonological segmentation, either based on syllabic organisation (like in French or in Dutch) or prosodic regularities (like in English or in Dutch) are thought of as useful to *help* in word recognition. The output of phonological organisation processes would therefore be available to directly influence lexical activation levels at later levels of processing (cf. Figure 1). According to this view –and admitting that the data provided by the phoneme detection task need

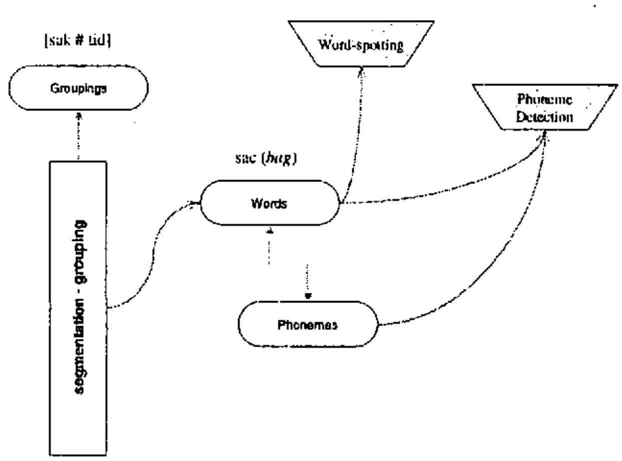

Figure 1: Direct influence of prelexical constraints on word-activation levels.

some kind of lexical identification– both tasks should lead to similar observations as, in each of these experimental paradigms, phonological segmentation would have to be taken into account by lexical levels of processing.

Another point of view may however be provided according to which lexical activation levels are independent from phonological organisation processes. We argue that phonological organisation provides a means to structuring the input without aiming to help in word identification. According to this proposal, facilitation occurs as a consequence of two independent sources of information [5]. Lexical activation levels account for the ability to *identify* a word in the speech input. The aim of phonological organisation processes is to establish groupings of segments in terms of higher order phonological structure. Combination of these sources of information may influence the ability to recognize a word in the speech stream. However, both would be rather independent. According to this view,

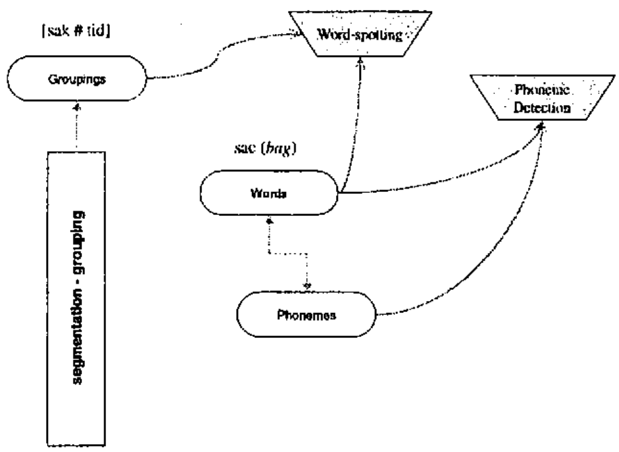

Figure 2: Common influence of prelexical constraints and word-activation levels on word identification.

phoneme detection data may be immune to the influence of phonological segmentation processes. Indeed, even if the decision made with respect to the presence of a target phoneme depends on lexical activation levels, speech structuring would not be able to influence observations in the phoneme detection task as

The only aim of organizing the speech input according to the phonological regularities which speakers have integrated during language acquisition would be to provide a means of structuring this acoustic stream. No direct influence of these processes on lexical stages would occur. Prelexical stages of phonological organisation would therefore constitute lexically–independent, general perceptual processes transfered in the domain of language processing. Similarly to the hierarchical organisation of a musical piece, or to the binding of unstructured elements in a visual scene, they would not be designed with a primarily *word recognition goal*. The following figure gives a visual insight into the validity of this hypothesis.

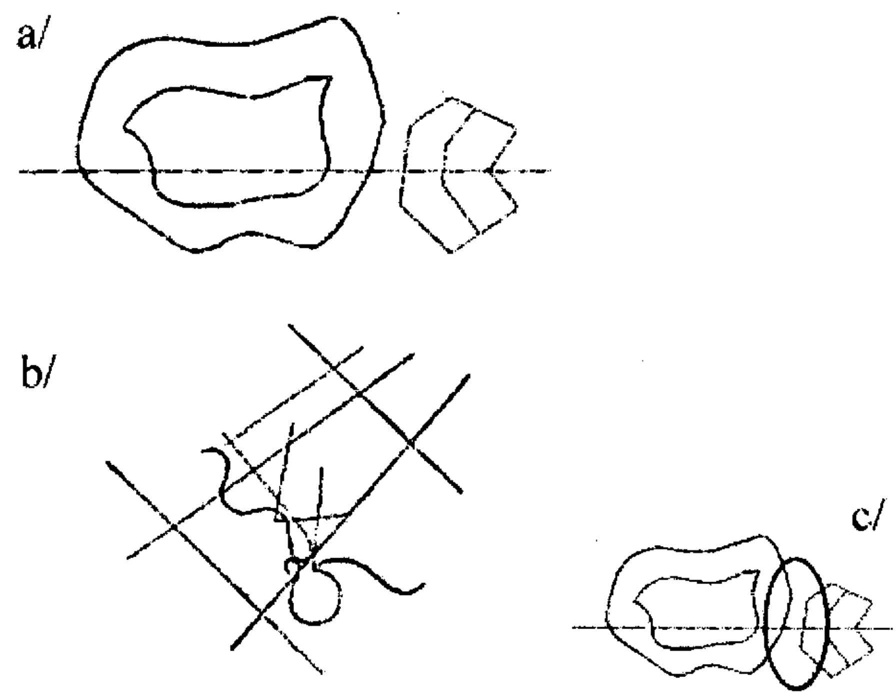

Figure 3: An illustration of structuring the input without modifying higher level processes. (a) No *known* structure pops-out from the scene. (b) It is clear that the figure '4' is embedded in the scene. Though it is also available in (a), we don't *see* it. This is a well known example of how perceptual organisation influences perception. There is no need for these processes to modify our internal representation of the figure '4' [4].

References

[1] A. Content, P. Mousty, and M. Radeau. Brulex, une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, 90:551–570, 1990.

[2] A. Cutler and D. Norris. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14:113–121, 1988.

[3] U.H. Frauenfelder, J. Segui, and T. Dijkstra. Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):77–91, 1990.

[4] W. Köhler. *Gestalt Psychology*. Horace Liveright, New-York, 1929.

[5] D. W. Massaro and M. M. Cohen. Phonological context in speech perception. *Perception & Psychophysics*, 34:338–348, 1983.

[6] J. M. McQueen. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39:21–46, 1998.

[7] J.M. McQueen. Word spotting. *Language and Cognitive Processes*, 11:695–699, 1996. In F. Grosjean & U. Frauenfelder (Eds.), A Guide to Spoken Word Recognition Paradigms (pp. 695-699). Hove, England: Psychology Press.

[8] J. Vroomen and B. de Gelder. Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory and Cognition*, 27(3):413–421, 1999.