WORD RECOGNITION: DO WE NEED PHONOLOGICAL REPRESENTATIONS?

Christophe Pallier
Laboratoire de Sciences Cognitives et Psycholinguistique
EHESS–CNRS, 54 bd Raspail, Paris, France

ABSTRACT

Under what format(s) are spoken words memorized by the brain? Are word forms stored as abstract phonological representations? Or rather, are they stored as detailed acoustic-phonetic representations? (For example as a set of acoustic exemplars associated with each word). We present a series of experiments whose results point to the existence of prelexical phonological processes in word recognition and suggest that spoken words are accessed using a phonological code.

1. INTRODUCTION

Linguistics makes a strong case for the psychological reality of phonological representations. It is important to assess how and when phonological representations are used by the brain. Some have argued that phonological representations may be used in speech production but not in speech perception [1, 2, 3]. They propose, instead, that word recognition involves a "direct" mapping from an acoustic representation of the input to the lexical representations. The series of experiments presented in this paper address whether phonological representations can be spared in speech perception.

2. LANGUAGE-SPECIFIC REPETITION PRIMING

The first experiment was designed to test whether the metrics of comparison between two word forms are universal or language-specific [4]. Indeed, theories postulating phonological prelexical processes naturally predict that an incoming spoken word is represented in a language-specific manner. However, one may argue that strong forms of the "direct acoustic access" hypothesis posit that word forms are represented and compared using a universal acoustic format.

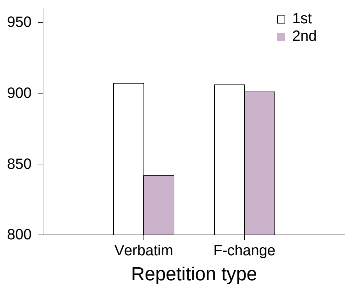
We tested two groups of people who spoke two dialects of the same language, that is they shared the same vocabulary, yet they had different phonemic systems. Actually, the first group of subjects consisted of native speakers of Catalan, while the second group consisted of Spanish speakers who were highly fluent in Catalan, as a result of a long and intensive exposure to Catalan starting around 5 years of age. In previous work [5], we demonstrated that, roughly speaking, those two groups do not have the same phonemic systems, as assessed by phonetic categorization and discrimination tasks.

The present experiment is based on the repetition effect, that is the fact that the processing of a stimulus is facilitated when it is repeated. Participants had to perform an auditory lexical decision task, with lists of words that contained some words repeated verbatim, and also some words repeated with one phonetic feature changed. Crucially, the phonetic feature alternation was phonemic for one population but not for the other. The question was whether the repetition effect would be depend on each subject's phonemic system or not.

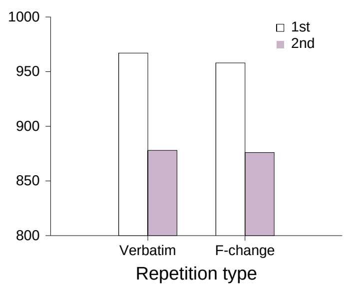
The results were unambiguous (see fig 1): the repetition effect is modulated by the phonemic system of the listener.

Figure 1: Average reaction times for the first and second occurrences of words, as a function of the relationship between the first and the second member of the pair ('same token' or 'feature change'). The repetition effect is measured by the difference in RT between the 2nd and the 1st occurrence.

Catalan-dominant:



Spanish-dominant:



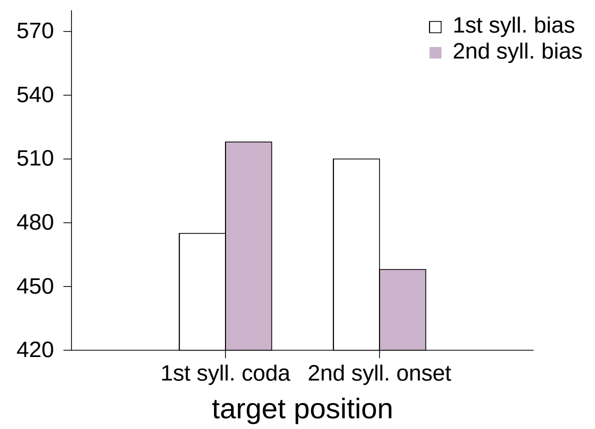
The fact that word recognition is language-specific is not necessarily problematic for direct access models: the acoustic records obviously depend on the history of the listeners and, therefore, on the language. If the Spanish have heard many “distorted” Catalan words, then it is not surprising that they have different representations for these words. It would be important to know whether or not they have heard a lot of “broken” Catalan. This seems very unlikely according to our Catalan informants. If one accepts that the two populations (Spanish and native Catalan) have been exposed to roughly similar samples of Catalan speech, this result then clearly poses a threat for direct acoustic access models with universal metrics of comparison.

3. ON-LINE ELABORATION OF THE SYLLABIC STRUCTURE

Phonological representations, as postulated by linguists, usually possess a hierarchical structure: a word is not just a simple concatenation of phonemes: these phoneme appear within a syllabic frame. For example, the word “caprice” can be represented as “[ka] [pris]” and the word “capture” as [kap] [ture]”. In both cases, [p] is the third phoneme, but it belongs to the first or to the second syllable, respectively.

The next experiment was designed to assess whether the perceptual system is sensitive to the syllabic position of phonemes [6]. Subjects had to perform a phoneme monitoring task in lists of words. In these lists, we had biased the probabilities that the target phoneme appeared either in the first or in the second syllable. The outcome, displayed in figure 2, revealed that the subjects had (implicitly) become “attuned” to the most probable syllabic position (see also [7, 8]).

Figure 2: Average phoneme detection times, as a function of the target’s position, of two groups of Ss induced to attend to coda and onset syllabic positions, respectively



Interestingly, other experiments have revealed that participants have become attuned, not to the whole syllable, but to a precise phoneme slot inside the syllabic structure [9, 10]. We also showed that it is not possible to train people to attend to phoneme position defined purely sequentially (e.g. to the third phoneme of utterances) [11]. These data suggest that the brain elaborates a syllabically structured representation of the stimuli, even in a task that does not require explicit manipulation of the syllable.[1]

These results, and the previous ones, lead to the notion that the perceptual system builds a language-specific representation that specifies the syllabic structure. We proposed that the syllabic structure plays the role of a corrector code: by imposing syllabic well-formedness constraints, the perceptual system may solve ambiguous or incomplete phonetic analyses. This idea was tested in the next experiment.

4. PHONOLOGICALLY INDUCED PERCEPTUAL CORRECTION

Japanese is a language whose phonology places strong constraints on admissible sequences of phonemes: most syllables are of CV type and the number of possible clusters of consonants is quite limited. If a Japanese speaker listens to an illegal sequence of consonants, this may trigger an “error signal” in his speech decoding system, which may then try to “correct” the stimulus to make it conform to the phonology of the language. Indeed, when asked to repeat a stimulus containing an illegal consonant cluster, Japanese speakers will typically insert a vowel “u” between the consonants, producing an utterance that respects the phonology of their language.

The task of repeating a stimulus involves both perceptual and production processes. Does the “correction” take place at the production stages, or is it performed at the perceptual stages? We examined this question by presenting Japanese listeners with illegal stimuli of the form VCCV, where the CC clusters were not admissible in Japanese [12]. More precisely, we built series of stimuli ranging from VCCV to VCuCV by increasing the duration of the middle vowel. The participants listened to the stimuli and had to report whether they heard a “u” in between the consonants or not. Their performances are plotted on figure 3, along with the results from a control group of native French speakers.

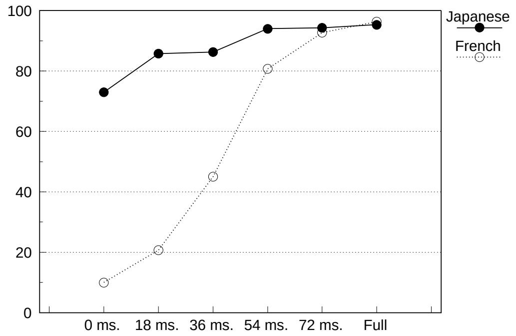
These data suggest that Japanese “hear” the vowel “u”, even when there is no acoustic correlates of it in the signal. This conclusion was strengthened by other experiments which used the ABX task: they showed that Japanese listeners had significant difficulties discriminating ”VCCV” from ”VCuCV” stimuli.

These results are compatible with the notion that the perceptual system tries to elaborate a phonological representation of incoming speech.

It is not immediately clear how direct access models could predict these results. Yet, one may try and argue that these effects result from the acoustic similarity between

1. One problem to note, however, with phoneme detection experiments, is that is it not clear to what extend they taps prelexical processes

Figure 3: Percentage of detection of a vowel “u” as a function of the duration of the “u” within VCuCV stimuli; comparison between native Japanese and native French listeners



the stimuli and items stored in the mental lexicon: If a stimulus like “ebzo” is presented and there are words like, for example, “ebuza”, “ibuzo”... in the mental lexicon of the Japanese listeners, these words may conspire to yield a “u” response.

This hypothesis was tested, and refuted, by Dupoux et al. [13]. We will give here only a simplified description of their experiment. They created pseudo words from real Japanese words by deleting the middle vowel which could be “u” or a “a”; e.g. `sokudo -> sokdo` and `mikado -> mikdo`. The main finding is that, in a lexical decision task, Japanese listeners made many false alarms on “u”-type pseudo-words but not on “a”-type ones; that is, they answered “word” to items like `sokdo`, but not to items like `mikdo`. If the acoustic distance between the stimulus and the items in the lexicon was all that mattered, there should have been no difference between the two types of pseudo-words. These results, however, are expected if the perceptual system inserts a “u” between two consonants at a prelexical stage.

5. CONCLUSION

Theories postulating that the brain elaborates phonological representations of spoken stimuli straightforwardly predict the outcomes of the experiments presented in this paper (see also [14]). It may be possible to tweak acoustic models to yield the same results (in particular, they are generally vague enough about the metrics of comparison between acoustic exemplars). We would certainly not claim that the debate between the two classes of theories is settled. For example, a wealth of data exists that suggest that sub-phonetic information can percolate up to the lexicon (e.g. [15]). To explain this, phonological models have to introduce the notion of graded phonological representations, and then, one must admit, the distinction between acoustic and phonological models starts to blur...

4. REFERENCES

[1] Klatt, D. H. (1981). Lexical representations for speech production and perception. In Myers, T., Laver, J., and Anderson, J. (Eds.), *The Cognitive Representation of Speech*, (pp. 11–31). North-Holland Publishing Company, Amsterdam.

[2] Pisoni, D. B. (1996). Some thoughts on “normalization” in speech perception. In Johnson, K. and Mullenix, J. W. (Eds.), *Talker Variability in Speech Processing*. Academic Press, San Diego.

[3] Goldinger, S. D. (1992). *Words and Voices: implicit and explicit memory for spoken words*. Ph.d., Indiana University.

[4] Pallier, C., Sebastián-Gallés, N., and Colomé, A. (1999). Phonological representations and repetition priming. In *Proceedings of Eurospeech ’99*, volume 4, (pp. 1907–1910).

[5] Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, **64**, 3, B9–B17.

[6] Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., and Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, **32**, 373–389.

[7] Finney, S., Protopapas, A., and Eimas, P. D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language*, **35**, 893–909.

[8] Pitt, M. A., Smith, K. L., and Klein, J. M. (1998). Syllabic effects in word processing: evidence from the structural induction paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 6, 1596–1611.

[9] Pitt, M. A. and Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language*, **29**, 611–632.

[10] Pallier, C. (1997). Phonemes and syllable in speech perception: size of the attentional focus in French. In *Proceedings of Eurospeech ’97*, volume 4, (pp. 2159–2162).

[11] Pallier, C. (1994). *Rôle de la syllabe dans la perception de la parole: études attentionnelles*. PhD thesis, Thèse de doctorat de l’École des hautes études en sciences sociales [available from the author].

[12] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 6, 1568–1578.

[13] Dupoux, E., Fushimi, T., Kakehi, K., and Mehler, J. (1999). Prelexical locus of an illusory vowel effect in japanese. In *Proceedings of Eurospeech ’99*, volume 4, (pp. 1675–1678).

[14] Marslen-Wilson, W., Nix, A., and Gaskell, M. G. (1995). Phonological variation in lexical access: Ab-

stractness, inference and English place assimilation. *Language and Cognitive Processes*, **10**, 3, 285–308.

[15] Andruski, J. E., Blumstein, S. E., and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, **52**, 3, 163–187.