PHONEMES AND SYLLABLES IN SPEECH PERCEPTION: SIZE OF ATTENTIONAL FOCUS IN FRENCH

Christophe Pallier
Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands &
Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-CNRS, Paris, France
E-mail: pallier@lscp.ehess.fr

ABSTRACT

A study by Pitt and Samuel (1990) found that English speakers could narrowly focus attention onto a precise phonemic position inside spoken words [1]. This led the authors to argue that the phoneme, rather than the syllable, is the primary unit of speech perception. Other evidence, obtained with a syllable detection paradigm, has been put forward to propose that the syllable is the unit of perception; yet, these experiments were ran with French speakers [2]. In the present study, we adapted Pitt & Samuel's phoneme detection experiment to French and found that French subjects behave exactly like English subjects: they too can focus attention on a precise phoneme. To explain both this result and the established sensitivity to the syllabic structure, we propose that the perceptual system automatically parses the speech signal into a syllabically-structured phonological representation.

1. INTRODUCTION

Both the phoneme and the syllable have often been proposed as the "primary unit of speech perception" [3-8]. Though this notion — which has rarely been clarified — has become less fashionable with the advent of "continuous" models of word recognition [e.g. 9], the controversy has generated many intriguing observations. One of these observations is the differential sensitivity of French and English listeners to the syllable [10]. The results of a series of studies, using syllable and phoneme detection tasks, were taken to indicate that French listeners, but not English listeners, use syllabic units to represent the speech signal [11, 12].

The evidence for the use of the syllable by French native speakers stems from the "syllable congruency effect" [2]: when asked to monitor for BA or BAL in words like BA#LANCE or BAL#CON (the '#' indicating the syllable boundary), French subjects are faster if the target precisely matches the first syllable of the stimulus. For example BAL is detected faster than BA in BAL#CON, while the opposite is observed for BA#LANCE. When English subjects are tested with the very same material, however, they do not display sensitivity to the syllable boundary [10, 13].

A very different approach has been used to suggest that the phoneme is the unit of perception for English speakers. One way to characterise a "perceptual unit" is to show that it can be the object of focal attention [14]. Following such a logic, Pitt & Samuel [1] tried to determine whether auditory attention could be focused on a unit as small as the phoneme. They had groups of subjects perform a generalised phoneme detection task with a list of CVCCVC words. In this experimental paradigm, subjects carry out a series of trials consisting of the presentation of a written phoneme followed by a spoken word; their task is to press a button, as quickly as possible, if they hear the target phoneme in the incoming word. Their reaction times are then measured. In Pitt & Samuel's study, the probability of occurrence of the target phoneme in the different consonant positions was manipulated: for four groups of subjects out of five, the target was more likely to occur in one consonant position than in the others (1st C for group 1, 2nd C for group 2...) . Thus, the subjects were effectively "conditioned" to expect the target in a precise sequential location. All groups shared some target-word pairs so that their detection times could be compared for all four consonant positions. The fifth group was a "control group" for which the four consonant positions had the same probability of occurrence. Each "experimental" group could be compared with this control. The results were clear: each group showed an advantage when the target occurred precisely at the location which was the most probable. Notably, this advantage did not extend to the other consonant inside the same syllable. That attention could be focused on the phoneme, the authors argued, demonstrated that the phoneme was in fact the unit of speech perception, at least for English speakers.

An important question in this study concerns the property used by the subjects to detect the regularity in the position of the target. It may be (as the authors seem to implicitly assume) that this property was the sequential position of the phoneme in question. However, a feature of the words used in Pitt & Samuel's study was that most had a "CVC#CVC" structure; thus subjects expecting, say, the third consonant, may in fact have been focusing on the onset of the second syllable; and subjects trained to detect the second consonant may have been focusing on the coda of the first syllable, and so on. Put differently, the relevant properties on which attention was focused may have been a position defined in terms of the syllabic structure rather than in terms of sequential phonemic position. This hypothesis was tested by Pallier et al. [15]. In their study, there were two groups of subjects: the first had to detect phonemes that occurred more often in the coda of the first syllable (e.g. P in caP#ture), and the second had to detect phonemes that occurred more often in the onset of the second syllable (e.g. P in ca#Price). The results proved that this manipulation indeed affected detection times on test words shared by the two groups: the first group was faster for codas of first syllables, and the second was faster for onsets of second syllables. This demonstrated that subjects could focus attention on a precise position in the syllabic structure of the stimuli [15].

The fact that syllabic structural position can be used does not rule out the possibility that sequential position cannot play a role. However, Pallier [16] showed that it was not possible to induce subjects to attend to a sequential position when the structural status of the target was varied. That is, subjects could not take advantage of the fact that a phoneme target was more often in the third sequential position when it was as likely to be a coda (caP#tif) than an onset (ca#Price). This fact demonstrates that sequential position is simply not a psychologically relevant property: subject do not automatically "count" phonemes.

Nevertheless, whereas Pitt and Samuel used American English listeners, the studies by Pallier et al. were conducted with French and Spanish subjects. In light of the previous studies concluding that French, but not English, relied on the syllable, one alternative explanation for the results is possible: French subjects may have been focusing attention on the whole syllable. The present study, an adaptation to French of the Pitt & Samuel study, was designed to assess this hypothesis.

2. EXPERIMENT

This experiment is a replication, with French subjects, of Pitt and Samuel's (1990) experiment 1: we compared four groups of Ss who were induced to attend respectively to the first, second, third and fourth consonantal position of CVC#CVC words.

2.1 Method

2.1.1 Material

One hundred and forty seven CVC#CVC French words were selected, providing four potential consonant locations for the targets (C1, C2, C3 and C4, corresponding to the sequential position inside the word). Four lists of 147 word-target phoneme trials were then constructed. The lists differed only in the target phonemes: the words and their order were the same in the four lists. Sixteen different types of phonemes were used for the targets. There were three categories of trials: Tests, Fillers and Foils. There were forty "Tests" trials: ten for each of the consonant location condition (C1, C2, C3 and C4). Care was taken that the phoneme sets were the same in the four conditions. Only the 70 Fillers distinguished the four lists: in the first list, the target phoneme was always the first consonant of the word; in the second list it was the second consonant, and so on. The 37 "Foil" trials were distracters for which the target did not occur in the word. The Test trials were always preceded by one or two Filler trials. Finally, we insured that two successive trials never contained the same phoneme target.

2.1.2 Procedure

The subjects were tested individually in a quiet room. They were seated in front a portable PC Toshiba 5200 that controlled the progress of the experiment (the stimuli were stored as 16 bit files on the computer and played back at 64 kHz by an OROS AU22 sound board). After reading the instructions explaining the generalised phoneme detection task, the subjects performed a short (unbiased) training with 7 trials. Then, the experiment proper started. The subjects were not informed that the phoneme target would occur more often in one location than in the other. Each trial began with the presentation of an upper case letter representing the phoneme target for 800 ms. The screen was then cleared, and one second later, a word was played in the headphones. The subject had to press a morse key when he detected the target (he had 2.5 sec from stimulus onset to answer). The reaction-time was measured from the target phoneme onset. The next trial started 5 sec after the beginning of the previous one, except when the subject had made a mistake (false alarm or miss). In such a case, a warning message was displayed for 2.5 second and the same trial was started again (only the first trial was considered for analysis, that is, as an error; this change in the original procedure aimed at diminishing the effects of errors on subsequent trials).

Several departures from Pitt & Samuel's original design are notable:

- in our case 75% (vs. 50% in the P&S study) of the words contained the target, and we used half as many test words, ending up with a lists of 147 stimuli rather than 480. This reduced the duration of the experiment from 45 minutes to 15 minutes.

- we used a go/no go paradigm rather than a yes/no decision. Our aim was to speed up response times and decrease error rates.

- we had only 4 groups of subjects and no control group. This is because a between-subjects design would have required many more subjects than the within-subject design that we planned for this experiment (see results section).

2.1.3 Subjects

Forty students from various universities in Paris participated in the experiment for which they received 20 FF. All were native speakers of French with no known auditory defect. Ten were randomly assigned to each of the four lists, yielding four groups.

2.2 Results

Mean detection times and error rates were computed for each subject. Table 1 displays the data averaged over groups and consonantal positions. Table 2 shows cost-benefit figures computed from the reaction time data by removing the mains effects of group and consonant position from each cell.

We performed two two-way analyses of variance on the mean reaction times, the first with subjects and the second with items as the random factor. The two factors were (a) (consonantal) "Position" (within-subjects and between- items) and (b) "Group" (between-subjects and within-items). Position yielded a significant effect ($F_1(3,108)=109$; $p<.001$ and $F_2(3,35)=20$; $p<.001$), as did the global interaction Group × Position ($F_1(9,108)=8.7$; $p<.001$ and $F_2(9,105)=9.6$; $p<.001$). We then examined, for all pairs of groups, the interactions Group × Position: All were significant at the .05 level both in the subject-based and in the item-based analysis.

Table 1. Mean reaction-times (in ms) and error-rates (in percents) of each group according to the consonantal position.

Group	Consonantal position							
	C1		C2		C3		C4	
Gr1	503	1	596	3	488	2	344	1
Gr2	494	0	436	1	439	1	353	7
Gr3	527	1	502	2	404	0	350	4
Gr4	640	0	590	3	505	1	290	1

Each cell is the mean of 100 measures (10 Ss × 10 items). The pooled standard deviation is 57.6 ms by subjects and 53.4 ms by items.

Table 2. Costs/Benefits computed from the reaction-times.

Group	Consonantal position			
	C1	C2	C3	C4
Gr1	-54*	49	13	-7
Gr2	-11	-59*	16	55
Gr3	7	-8	-34*	36
Gr4	59	19	6	-84*

Each cell is obtained from the corresponding one in table 1 by computing ($y_{ij}= x_{ij}-x_{i.}-x_{.j}+x_{..}$). (*= expected position)

2.3 Discussion

First and foremost, each group had the largest benefit when the phoneme target occurred in its expected position (see diagonal of Table 2). This was statistically significant as attested by the significance of all two by two interactions. We thus reproduce the main result of Pitt & Samuel, but this time with French subjects.

Second, our subjects were faster (460 ms) than Pitt & Samuel's (701 ms) and the error rates were quite low (false alarms: 2.7%; missed targets: 1.3%). Three differences between their study and ours can be invoked to try to explain the difference: (a) we used a go/no-go paradigm rather than yes/no decision (b) we gave feedback on errors (c) our experiment was much shorter (about 15 min) than theirs (about 40 min).

Finally, we observe that the reaction times substantially decrease as the position of the target phoneme approaches the end of the word. We think that several factors may conspire to bring this about: (a) coarticulation information before the target phoneme can help the response (b) with position, the potential influence of lexical knowledge increases [17, 18] (c) subjects response "threshold" may decrease with time (It is a general fact that detection time decreases as the serial position of the target in a list of stimuli increases (see ref. [19]) (d) the end of the stimulus can act as a go signal.

3. CONCLUSION

This experiment replicates with French subjects the pattern of results previously obtained with American subjects: French listeners too can focus their attention on a unit as small as a phoneme. This result adds to the ones reported in [15, 20]: together they demonstrate that listeners can focus attention on phoneme-sized units whose position is defined in terms of the syllabic structure of the stimulus.

Most models of speech perception make the assumption that the brain extracts a linear string of units (phonemes, syllables, etc.) from the speech signal. The data presented here are better interpreted by supposing that the speech processing system elaborates, in real-time, a hierarchical, syllabically-structured representation of the stimulus [16].

4. ACKNOWLEDGEMENTS

The author was supported by grants from the Direction Générale des Études et Techniques and from the Fyssen Foundation.

5. REFERENCES

[1] M. A. Pitt and A. G. Samuel, “Attentional Allocation during Speech Perception: How fine is the focus?,” *Journal of Memory and Language*, vol. 29, pp. 611-632, 1990.

[2] J. Mehler, J. Y. Dommergues, U. Frauenfelder, and J. Segui, “The syllable's role in speech segmentation,” *Journal of Verbal Learning and Verbal Behavior*, vol. 20, pp. 298-305, 1981.

[3] H. Savin and T. Bever, “The nonperceptual reality of the phoneme,” *Journal of Verbal Learning and Verbal Behavior*, vol. 9, pp. 295-302, 1970.

[4] D. W. Massaro, “Perceptual units in Speech Recognition,” *JEP*, vol. 102, pp. 199-208, 1974.

[5] J. Mehler, “The role of syllables in speech processing: Infant and adult data,” *Philosophical Transactions of the Royal Society*, vol. 295, pp. 333-352, 1981.

[6] E. Dupoux, “Prelexical processing: the syllabic hypothesis revisited,” in *Cognitive models of speech processing: The second sperlonga meeting*, G. T. M. Altmann and R. Shillcock, Eds. Hove East Sussex UK: LEA, 1993, pp. 81-114.

[7] S. Decoene, “Testing the speech unit hypothesis with the primed matching task: phoneme categories are perceptually basic,” *Perception & Psychophysics*, vol. 53, pp. 601-616, 1993.

[8] J. Segui, “The syllable: A basic perceptual unit in speech perception ?,” in *Attention and Performance X*, H. Bouma and D. G. Bouwhuis, Eds. Hillsdale NJ: Erlbaum, 1984, pp. 165-181.

[9] J. L. McClelland and J. L. Elman, “The TRACE model of speech perception,” *Cognitive Psychology*, vol. 18, pp. 1-86, 1986.

[10] A. Cutler, J. Mehler, D. Norris, and J. Segui, “The syllable's differing role in the segmentation of French and English,” *Journal of Memory and Language*, vol. 25, pp. 385-400, 1986.

[11] D. G. Norris and A. Cutler, “The relative accessibility of phonemes and syllables,” *Perception & Psychophysics*, vol. 43, pp. 541-550, 1988.

[12] J. Segui, E. Dupoux, and J. Mehler, “The role of the syllable in speech segmentation, phoneme identification and lexical access,” in *Cognitive models of speech processing: psycholinguistic and computational perspectives*, G. T. M. Altmann, Ed. Cambridge Mass.: MIT Press, 1990, pp. 263-280.

[13] A. Cutler, J. Mehler, D. Norris, and J. Segui, “A language specific comprehension strategy,” *Nature*, vol. 304, pp. 159-160, 1983.

[14] U. Neisser, *Cognitive Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, 1967.

[15] C. Pallier, N. Sebastian-Gallés, T. Felguera, A. Christophe, and J. Mehler, “Attentional allocation within syllabic structure of spoken words,” *Journal of Memory and Language*, vol. 32, pp. 373-389, 1993.

[16] C. Pallier, “Rôle de la syllabe dans la perception de la parole: études attentionnelles,” Paris: Ecole des Hautes Etudes en Sciences Sociales, 1994 [available from the author].

[17] U. H. Frauenfelder, J. Segui, and T. Dijkstra, “Lexical Effects in Phonemic Processing: Facilitatory or Inhibitory ?,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16, pp. 77-91, 1990.

[18] U. H. Frauenfelder and J. Segui, “Phoneme monitoring and lexical processing: Evidence for associative context effects,” *Memory and Cognition*, vol. 17, pp. 134-140, 1989.

[19] R. D. Luce, *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press, 1986.

[20] S. A. Finney, A. Protopapas, and P. D. Eimas, “Attentional Allocation to Syllables in American English,” *Journal of Memory and Language*, vol. 35, pp. 893-909, 1996.