British Journal of Psychology (1981), 72, 471–477 *Printed in Great Britain*

Phoneme monitoring, syllable monitoring and lexical access

J. Segui, U. Frauenfelder and **J. Mehler**

In this experiment we investigated the role of the lexical status (word/non-word) of the target item in the determination of phoneme and syllable detection times. Subjects monitored either the initial stop consonants (/b/, /d/, /p/) or the initial CV syllables (/ba/, /de/, /pi/) in bisyllabic target items (word/non-word) in mixed lists. The lexical status of the target item did not introduce significant differences for phoneme or syllable detection times. However, significant differences were found between the phoneme and syllable detection times. In addition, a strong correlation between the phoneme and syllable RTs for each item (word/non-word) was obtained. The first result shows that subjects can respond to both initial phonemes and syllables prior to lexical access. The second result suggests that phoneme detection is highly dependent on syllable identification.

According to a traditional point of view, the perception of speech is seen as a series of processing stages during which smaller units are combined to create larger ones. This bottom-up perspective assumes that there are in the signal invariant acoustic cues corresponding to the smallest perceptual unit, the phoneme. The failure to discover these acoustic invariants has led certain investigators to question the validity of this theoretical perspective. Savin & Bever (1970), on the basis of observed differences between item initial phoneme and syllable detection times, concluded that syllable recognition precedes phoneme recognition. More precisely, the authors suggest that the phoneme is not directly perceived but is derived from an analysis of the syllable. The hypothesis according to which the recognition of higher units precedes that of immediately lower level units was generalized by Foss & Swinney (1973) to the identification of other linguistic units. Thus, they state, 'smaller units are identified by fractionating larger ones' (p. 254). This hypothesis leads to the prediction that phonemes and syllables are identified by fractionating words. The results obtained by Morton & Long (1976) support this view. Indeed, Morton & Long demonstrated a relationship between phoneme detection times and the transitional probability of the target words. In another study (Rubin *et al.*, 1976), subjects were presented sequences of monosyllabic items (words or non-words) and were asked to detect either of the two following item initial phonemes: /b/ or /s/. When the subjects received mixed sequences of words and non-words, the detection times for phonemes in words were significantly shorter than those in non-words. The word facilitation effect for phoneme detection led the authors to conclude that lexical access precedes phoneme detection.

More recently, however, Foss & Blank (1980) have reported a series of phoneme monitoring results within sentences that are inconsistent with those obtained by Rubin *et al.* (1976). In one experiment the lexical status (word/non-word) of either the target-bearing item or the item preceding the target was manipulated. The results indicated that the lexical status of the pre-target item, the item immediately preceding the target, and *not* that of the target item itself affected the detection times. In a second experiment, the frequency of the pre-target and target words varied. Similarly, the RTs were affected only by the frequency of the pre-target word. The important role that the pre-target word plays in the determination of phoneme detection times has been found in other studies (Mehler *et al.*, 1978; Newman & Dell, 1978). The absence of an effect of either the frequency or the lexical status of the target item was taken by the authors to show that phonemes are generally detected on the basis of a bottom-up or (prelexical access) 'phonetic code'.

0007-1269/81/040471-07 $02.00/0

However, according to their 'dual code' hypothesis (which resembles the race models proposed in Newman & Dell, 1978, and in Cutler & Norris, 1979), phonemes can also be identified on the basis of a (postlexical access) 'phonological code' when they are in words which are highly predictable from the linguistic context.

While the 'dual code' hypothesis can account for the results obtained by Morton & Long (1976), it would not predict those obtained by Rubin *et al.* (1976), since the target words were in sequences where there was no predictive context. However, Foss & Blank (1980) offer an alternative explanation for these results in terms of the unusual task used: subjects were required to make a choice response (a different response key for each phoneme). The complexity of this task led to an increase in the momentary processing load, thereby forcing the subject to use the postlexical access 'phonological code' rather than the more ephemeral 'phonetic code'. Another potential problem of the studies by Rubin *et al.* (1976) concerns the use of monosyllabic items, which makes differentiating between syllabic and lexical access impossible.

The distinction between syllabic and lexical levels is critical for testing the principal hypothesis advanced here, according to which both item-initial phonemes and syllables are generally detected prior to lexical access. While this view is inconsistent with the generalized hypothesis of Foss & Swinney (1973), it is in accord with the 'dual code' hypothesis.

Our secondary hypothesis which holds that phoneme recognition depends on prior syllable identification is in disagreement with the 'dual code' hypothesis. According to a recent version of this hypothesis (Foss *et al.*, 1980), the representation of the speech signal is computed in terms of phoneme-sized units via strictly bottom-up processes.

However, in a recent study (Mehler *et al.*, in press *a*) it was shown that the time to detect a sequence of phonemic segments depends on the syllabic status of this sequence in the target word. Faster reaction times are observed when the sequence corresponds to a syllable in the word. For instance, the sequence /pa/ is detected faster than the sequence /pal/ in the word 'pa-lace'. The inverse result was obtained for the same targets in the word 'pal-mier'. These results suggest that the speech signal is computed in terms of syllable-sized units. Thus, syllables may be processed via a bottom-up analysis, whereas phonemes are derived from syllabic units. In our perspective the syllable constitutes the major point of departure from which phonemes can be identified (if the task so requires) and from which the higher level lexical units can be accessed (Mehler *et al.*, in press *b*).

In the experiment presented here, bisyllabic items whose lexical status is determined by the second syllable were used. With bisyllabic items the possible effect of lexical status on detection time can be measured not only for the initial phonemes but also for the first syllable.

Method

Subjects

Twenty adults in the Parisian university community served as subjects. Ten were assigned to the first list (phoneme monitoring) and the remaining 10 received the second list (syllable monitoring).

Materials and apparatus

Six monomorphemic bisyllabic words of $C_1V_1C_2V_2(C_3)$ structure were selected for each of the three following C_1V_1 syllables: /pi/, /ba/, /de/. For each of these 18 words of approximately the same frequency, a matching non-word respecting French phonological rules was created by altering C_2 by one or two distinctive features (generally voicing). Sample items and targets are presented in Table 1.

The targets (both words and non-words) were included as the last item in 36 experimental sequences with the word and matching non-word always in the same position in their respective sequences. This position varied from 2nd to 5th. Thus, the experimental sequences were composed of

Table 1. Sample items and targets

Target initial phoneme	Target initial syllable	Word	Non-word
/d/	/de/	décor	dégor
/b/	/ba/	bateau	bapeau
/p/	/pi/	pilon	piron

from two to five items, that is, one to four filler items plus the target item. The filler items for any given sequence were an approximately equal number of mono- , bi- , and trisyllabic words and non-words. Another 3 experimental sequences each containing one of the monosyllabic targets were constructed with similar fillers. In addition 21 distractor sequences (constructed in the same way as the experimental ones) were produced. In these sequences the position of the target item varied from 1 to 6 with some sequences that had no target at all. These distractor sequences were used to prevent the subject from anticipating target positions. No sequences contained more than one target item. The initial phoneme of the pre-target and target items always differed by more than three distinctive features. The vowel environments (V1) for certain distractor target items were different from those of the experimental target items (mismatch condition; pa..., be..., pé...).

The 60 sequences were arranged in random order along with 8 warm-up sequences into a list. This list was recorded by a female native speaker of French at a normal rate using a two-track Ampex AG 440B tape-recorder at 15 in/s. Items within a sequence were separated by 2 s intervals while sequences were separated by 5 s. The entire set of materials is listed in the Appendix.

On the two prepared tapes each sequence was preceded by a target specification: for the phoneme list (List I) a place whose name begins with the target phoneme was pronounced. The place name was chosen such that it always began with a consonant cluster to prevent the subject from extracting the first syllable of the place name: /b/ as in 'Bretagne'. For the syllable list (List II) the targets were also specified with a place name: /ba/ as in 'Babylone'. These two separate lists were used in order to avoid presenting the same subject with an identical item for each target type.

The subjects' RTs were measured by the electronic clock in a PDP12 which was triggered by a click placed on the inaudible channel of the tape. Clicks were aligned manually by running the tape back and forth across the playback head. A correction for each click was then measured by means of a two-channel oscilloscope and edited into the data collection program.

Procedure

Subjects received written instructions explaining that they were going to hear a series of sequences made up of a varying number of real and nonsense words. Subjects were instructed to respond to the items beginning with the target specified before each sequence with a place name (/b/ as in 'Bretagne') by pressing a key as fast as possible. Subjects, tested individually or in pairs, received the experimental tape during sessions that lasted about 20 min.

Results

The mean reaction times for each subject and for each item were computed. Reaction times which were shorter than 100 ms and longer than 1000 ms were omitted from the calculation of the means. The omitted data made up 4·0 per cent of the data for phonemes and 4·6 per cent for the syllables. No difference in the number of omissions for words and non-words was found on either tape.

The overall means found in Table 2 show that the lexical status (word or non-word) of the target item did not lead to reaction-time differences for either phonemes or syllables.

A two-way analysis of variance taking the subjects as the random variable and the target type as the inter-subject factor and lexical status as the intra-subject factor was conducted. Only the first factor introduced a marginally significant difference, $F = 3{\cdot}55$, d.f. = 1, 18, $P < 0{\cdot}10$ ($F < 1$ for lexical status and interaction). Another analysis of variance taking the

Table 2. Mean RT (ms) as a function of lexical status (word/non-word and target type (phoneme/syllable)

	Word	Non-word
Phoneme	347	346
Syllable	285	281

items as random variables gave $F = 55{\cdot}26$, d.f. $= 1, 17$, $P < 0{\cdot}001$ for the target-type factor ($F < 1$ for lexical status factor and interaction).

The weakly significant effect found in the first analysis of variance is certainly related to the relatively small number of subjects in the two independent groups. Given the total absence of interaction a one-tailed t test for independent groups was performed giving a significant result ($t = 1{\cdot}8$, d.f. $= 18$, $P < 0{\cdot}05$). It is important to note that the mean differences between phoneme and syllable detection times (62 ms for words and 65 ms for non-words) are comparable to those obtained by other investigators (Savin & Bever, 1970; Foss & Swinney, 1973).

Table 3 presents the results for the three categories of items used. While there is a definite effect of the target type (phoneme vs. syllable), no such effect is found for the lexical status of the target item.

Table 3. Mean RT (ms) as a function of lexical status, target type and item category

	Item category					
	/de/		/ba/		/pi/	
	Word	Non-word	Word	Non-word	Word	Non-word
Phoneme	312	314	348	332	381	393
Syllable	264	270	285	276	306	297

Table 3 also shows the existence of a correlation between the phoneme and syllable RTs for the different categories: RT: /de/ < /ba/ < /pi/. In order to determine the strength of this correlation between phoneme and syllable reaction times a Bravais–Pearson correlation coefficient was calculated for all the 36 items used: (18 words and 18 non-words), $r = 0{\cdot}62$, $P < 0{\cdot}01$. The correlation was significant for the words ($r = 0{\cdot}54$, $P < 0{\cdot}02$) and for the non-words ($r = 0{\cdot}68$, $P < 0{\cdot}01$).

Table 4. Mean RT (ms) for monosyllabic target items as a function of item category and target type

	Item category		
	/de/	/ba/	/pi/
Phoneme	331	348	426
Syllable	261	293	339
	70	55	87

The results for monosyllabic target items (shown in Table 4) reveal the same tendencies as those obtained when these syllables appear in bisyllabic items. Indeed, the same order between the RTs for the different categories of items (/de/ < /ba/ < /pi/) was obtained as well as an effect of target type (RT syllable < RT phoneme).

Discussion

The results of this experiment clearly show that the detection of both phonemes and syllables does not depend on the lexical status of the bisyllabic item in which they are found. Thus, initial phonemes and syllables can be detected before lexical access. These results are consonant with those obtained by Foss & Blank (1980) and can be explained in terms of the 'dual code' hypothesis. This hypothesis assumes that, in the absence of predictive context, listeners can respond to word initial phonemes on the basis of a prelexical (phonetic) code.

However, the fact that syllables are detected more rapidly than phonemes appears to be inconsistent with this purely bottom-up account. Indeed, it has been claimed that the syllable is perceived (Savin & Bever, 1970) or identified (Foss & Swinney, 1973) before the phoneme and that identification of the latter depends on analysis of the former. This line of reasoning has received a considerable amount of criticism by investigators who have proposed alternative explanations. Healy & Cutting (1976) have insisted on differences in identifiability of the different units. McNeill & Lindig (1973) interpreted the difference between phoneme and syllable detection times to be the result of the distance in linguistic level between the target and the search list. Finally, in two recent studies, this difference has been attributed either to the uncertainty concerning the syllabic context of the phoneme target (Swinney & Prather, 1980) or to the mismatch between the target and the stimulus (Mills, 1980).

However, there are some serious problems with these studies, as has been pointed out elsewhere (Mehler *et al.*, in press *b*). Indeed, the experiments using a target-stimulus match with constant vowel environment (McNeill & Lindig, 1973; Swinney & Prather, 1980; Mills, 1980) confuse syllable and phoneme monitoring. The subjects most probably make use of the same syllabic target representation for both conditions even if from an experimental point of view different targets are given. Therefore, a target–stimulus mismatch condition as found in our experiment constitutes a necessary condition for comparing phoneme and syllable detection times.

In conclusion, the results obtained here indicate that the recognition of the first phoneme or syllable in a word does not require prior access to that word. Furthermore, the recognition of a phoneme seems to depend on the previous recognition of the syllable to which it belongs. This interpretation gains support from a recent study by Blumstein & Stevens (1980) who, using brief synthetic stimuli as short as 10–40 ms sampled from the onset of the CV syllable, demonstrated that these stimuli generally contained sufficient information for both C and V identification. This result is taken by the authors to show that 'the brief stimulus signals the identity of the *syllable*, which is processed by the listener as a unitary percept on a single event. Having identified the syllable, the listener is then able to indicate its consonantal and vocalic components' (1980, p. 660). On the basis of these results and others recently obtained on lexical access (Cole & Jakimik, 1980; Cole *et al.*, 1980) it appears that the syllable plays a critical role in the perception of speech. Indeed, in accordance with our hypotheses, the syllable can be seen as the structural unit from which both higher and lower level analyses originate.

Acknowledgements

This work was supported by grants from the Harry Frank Guggenheim Foundation (1977–1978) to J. Mehler and the Délégation Generale à la Recherche Scientifique no. 807334.

References

Blumstein, S. E. & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Accoustical Society of America*, **67**(2), 648–662.

Cole, R. A. & Jakimik, J. (1980): How are syllables used to recognize words? *Journal of the Acoustical Society of America*, **67**(3), 965–970.

Cole, R. A., Jakimik, J. & Cooper, W. E. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America*, **67**(4), 1323–1332.

Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (eds), *Sentence Processing: Psycholinguistic Studies Presented to Merill Garrett*. Hillsdale, NJ: Erlbaum.

Foss, D. J. & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology*, **12**, 1–31.

Foss, D. J., Harwood, D. A. & Blank, M. A. (1980). Deciphering decoding decisions data and devices. In R. A. Cole (ed.), *Perception and Production of Fluent Speech*. Hillsdale, NJ: Erlbaum.

Foss, D. J. & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, **12**, 246–257.

Healy, A. F. & Cutting, J. E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, **15**, 73–83.

McNeill, D. & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, **12**, 419–430.

Mehler, J., Dommergues, J. Y., Frauenfelder, U. & Segui, (in press *a*). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*.

Mehler, J., Segui, J. & Carey, P. (1978). Tails of words: Monitoring ambiguity. *Journal of Verbal Learning and Verbal Behavior*, **17**, 29–37.

Mehler, J., Segui, J. & Frauenfelder, U. (in press *b*). The role of the syllable in language acquisition and perception. In T. F. Myers, J. Laver & J. Anderson (eds), *The Cognitive Representation of Speech*, Advances in Psychology series. Amsterdam/New York: North-Holland (in press).

Mills, C. A. (1980). Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior*, **19**, 75–83.

Morton, J. & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, **15**, 43–52.

Newman, J. E. & Dell, G. S. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior*, **17**, 359–374.

Rubin, P., Turvey, M. T. & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, **19**(5), 394–398.

Savin, H. B. & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, **9**, 295–302.

Swinney, D. A. & Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, **27**(2), 104–110.

Received 16 October 1980; revised version received 5 February 1981

Requests for reprints should be addressed to J. Segui, Laboratoire de Psychologie Expérimentale, 28 rue serpente, 75006 Paris, France.

U. Frauenfelder and J. Mehler are at the Centre d'Etude des Processus Cognitifs et du Langage, CNRS.

Appendix

Linguistic materials—word and non-word targets

Word	*Non-word*	*Word*	*Non-word*
débat	dépat	baquet	badet
dépot	débot	basset	bafet
décor	dégor	bavard	bafard
débit	dégui	piquet	piget
délit	dérit	pichet	pifet
défi	déchit	pipeau	piteau
badaud	bacaud	pivot	pifot
bateau	bapeau	piton	pidon
bassin	bafin	pilon	piron

Experimental sequences

pi/*chalo* aubergine kis sol pito
dé/*nolite* général *sul palivo* car dément
dé/estomac *til orneci* dépat
ba/*rapat* amérique *lib* dètail *ordina* bataille
ba/fou *cof* orphelin basset
pi/diner *badolé* car léopard *pécour fone*
dé/esturé défi
pi/pilaf
ba/met *fatu* pavillon *termipo* ti batard
dé/glaçon ferogé clachot mare débit
ba/bavard
pi/arme *isorage sou asticot ticon pim*
pi/*pitaf*
pi/*kra pifet*
ba/savon *bachin*
dé/ile iglan dépot
ba/poème *essan* créature gouf ba
ba/noir *bafar*
ba/cheminot *frum badé*
dé/déguir
ba/folle *madon* bateau
ba/mer *mouba* agricole *cacabri* maison *bamar*
pi/effort *esturé* pi
ba/chimerique *fram baquet*
ba/*imo* nuque langouste *lompurie bako*
dé/climat *alomé toili* mer *dégui*
dé/turbulent *chon armico* débat
ba/cadeau *barul*
ba/*bamar*
pi/sens charité *armi* pivot
dé/*moc* salade *dégor*
pi/arbre *térogué* cou baguette *tilon* pile
ba/nez *mitor bapo*
pi/*soun* armée piton
dé/sabotage *cirul* château dé
ba/*dug* écarté houx *binar* trier volinar
pi/branco adomé pigeon
ba/rue *clon* olivier *bafet*
dé/*écroufar* clou iglos *débo*
ba/*irdan* mule linguiste *linfitar* badeau
dé/désir
pi/but *edo chamiro* amitié piquet
pi/*nol* carré *pidon*
dé/*ousate* exemplaire *pem* fou careau *deran*
pi/*gouf pichot*
pi/*sarou* sourd carabine *cursilé* pilon
pi/mal *acar* assuré amateur *piguet*
dé/*muc* secrétaire décor
dé/suite *falar kis* divan *baroufis* tir
dé/*ermeni déchit*
pi/pâté *bolur* théoreme tasse *salito cra*
pi/sot chocolat *arpou pifot*
de/*orma* oiseau *déri*