

CHAPTER 4

Unifying Awareness and On-Line Studies of Speech: A tentative Framework

Emmanuel Dupoux
France Telecom, Paris &
Laboratoire de Science Cognitive
et Psycholinguistique, CNRS-EHESS
Paris - France

Jacques Mehler
Laboratoire de Science Cognitive
et Psycholinguistique, CNRS-EHESS
Paris - France

ABSTRACT

Generally, studies of speech recognition are related to theories of performance while studies of awareness are thought to bear upon language competence. In our conception, both areas of research contribute to our understanding of processing and of the representations that the subjects use when listening to speech. We present a unitary framework within which it becomes possible to incorporate the results from on-line speech recognition studies and from studies of the awareness that the language user has of speech segments. In particular, we argue that it is necessary to include a description of the manner in which acoustic-phonetic information is transduced, and represented in order for us to understand how subjects come to decide to respond or not in a psycholinguistic experiment. Particular attention is given to the data from on-line chunk detection experiments and to the potential role of orthographic representation.

As we all know, learning to speak and learning to read are very different processes. While one acquires speech by mere exposure, one learns to read the hard way, namely, at school. Not only is schooling required, but many years of instruction are needed to turn illiterates into skilled readers. Nonetheless, we have always been fascinated by the observation that the very young child seems to be *unaware* that one can view speech as a sequence of phonemes.

Research initiated by Liberman and her colleagues (Liberman, Shankweiler, Fischer, & Carter, 1974), but greatly expanded by a team of Belgian psycholinguists led by P. Bertelson (see the collection of papers edited by Bertelson, 1986), suggests that phonemic *awareness* arises while children are learning to read thanks to the conversion of graphemes into phonemes (Morais, Cary, Alegria, & Bertelson, 1979). This finding has received more attention from educational psychologists than from psycholinguists. This is surprising since, regardless of whether they study speech perception or reading, psycholinguists relate their findings to the units used to process speech, see Savin and Bever (1970), Spoehr (1981), Taft and Forster (1976),

and Mehler (1981), among many others. We should acknowledge, from the beginning of this chapter, that our intuitions lead us to expect that metalinguistic competence as it is explored in awareness studies reflects important properties of the speech perception apparatus.

Tacitly, psycholinguists have often behaved as if the study of metalinguistic competence and that of on-line speech perception are quite unrelated. Metalinguistic studies supposedly reflect the knowledge of the language while on-line studies, at best, reflect the subjects' encapsulated processes (see Pylyshyn, 1984). Though this attitude, to the best of our knowledge, has never been explicitly defended, observing how the fields have grown shows that the above assertion describes an underlying reality. Our belief is that this behaviour is profoundly misguided. It is not that we are willing to give up the distinction between competence and performance or between knowledge and encapsulated processes. Rather, we believe that the methods used in studies of awareness and on-line speech perception rely on representations brought about by the encapsulated processing system. In order to play metalinguistic games the subject has to use some representation or another. Whichever representation he/she might use it must be one that is computed by the processing system. Likewise, in chronometric studies in order to perform speeded tasks, and even to memorize targets that have to be detected, subjects have to rely on some representations. These representations might be phonetic, phonemic or even orthographic, and thus reflect the knowledge of language that the subjects have attained, just as in the studies of awareness.

The purpose we pursue here is to try and present a preliminary framework to illustrate the reciprocal advantage one might draw by formulating a unitary model within which both awareness and chronometric studies can be couched. The model, which will be detailed below, suggests that subjects behave on the basis of an encapsulated information processing system, a representational system of the surrounding information, and a non-modular decision mechanism. An important consequence of adopting such a unitary model is that psycholinguists will have to include considerations about literacy in the interpretation of on-line studies. Indeed, literacy adds representations which make it possible for subjects to perform in tasks like phoneme-monitoring experiment which they would otherwise be unable to do. Given this observation, it is surprising that psycholinguists have neglected the potential effects of cultural representations. A few exceptions can be mentioned. For instance, the ability to detect rhymes in spoken words is greatly affected by orthographic similarity (Seidenberg & Tanenhaus, 1979; see also Tanenhaus, Flanigan, & Seidenberg, 1980). Likewise, Cutler, Norris, and van Ooyen (1990) speculate that the behaviour of subjects in a vowel detection task may be influenced by orthography (see also Taft & Hambly, 1985).

Psycholinguists acknowledge that literacy adds new representations to the existing ones. In our framework, we have to say that these new representations can have a potentially large impact on subjects' performance in

psycholinguistic experiments. With these facts in mind, it becomes possible to propose some tentative but intriguing alternative accounts to established discoveries. Our attempt has to be looked upon as an *exercice de style* rather than as a definitive proposal. Though we remain uncertain about the details we present, we are persuaded that some version of our proposal will eventually be correct. But before we return to the larger theoretical issues, we have to untangle the appallingly prosaic problems of methodology. Indeed, our framework requires us to model not only the processing devices involved in speech perception, but also the ways in which tasks can be performed by subjects. Thus, the problems of methods and procedures, far from being prosaic, become central. In the next section we explore psycholinguistic methods and present a detailed version of our model of the subject's behaviour.

HOW ARE PSYCHOLINGUISTIC EXPERIMENTS CARRIED OUT?

Although psycholinguistics is a rather old discipline, its experimental history is not that long. Indeed, it is only in the last thirty years or so that most of the experimental work in the area has taken place, and it has taken place mostly in the domain of reading. The advent of digital speech editors however, has changed this since it has made it possible to control auditory stimulus presentation as easily as in the visual modality. In most acoustic-phonetic studies subjects had to classify or discriminate stimuli that varied along one or several acoustic dimensions. The psycholinguistic study of speech has added another dependent variable: reaction times. In these studies experimenters generally assume that subjects' latencies reflect processing complexity, more or less directly. In one of the first chronometric experiments in the area of speech, Savin and Bever (1970) found that subjects tend to respond faster to syllables than to their first phonemes when time is measured from the beginning of the stimulus. This difference in latency was interpreted by the authors as suggesting that the syllable is the unit or component that has processing reality while the phoneme, at least for perception, is a derived construct without any processing reality. Regardless of whether the authors were right or wrong, the rationale they used is by and large one that we have all relied upon ever since.

However, it is blatant that latencies individuate complex underlying processes and it seems unlikely that reaction times can uncover processing priorities in the absence of a model of the subjects' behaviour in a given situation. Experimental data is always coloured by the way in which it was gathered. According to the method that is employed, subjects execute one response or another and use criteria that depend on many factors to effect their responses. Even when subjects are told to press a response key whenever they detect a phoneme, or a longer segment, like the syllable, they still may use a variety of criteria to verify that they have or have not heard the target. They may also entertain different hypotheses about the experimenter's motivations

for running the experiment. It is known that subjects' behaviour depends on the pay-off matrix under which they operate and it is easy to demonstrate that there is a speed accuracy trade off that affects most experiments of this kind (for reviews see Luce, 1986; Parasuraman, 1986). The same is true when subjects are instructed to detect a change in speaker, a switch in ear of presentation, or an artifact like a click, regardless of whether the response is a two way classification or of a go-no go sort. In all these cases, the signal is processed and a representation is retrieved. Even so, a response is made only after certain verifications have been effected. A complex cognitive system seems to scan the internal state of the organism and the response buffer is triggered, if and only if, the representation being looked for obtains and the pertinent verifications give satisfaction. Only then is the response triggered. These criteria may be very different for the different tasks.

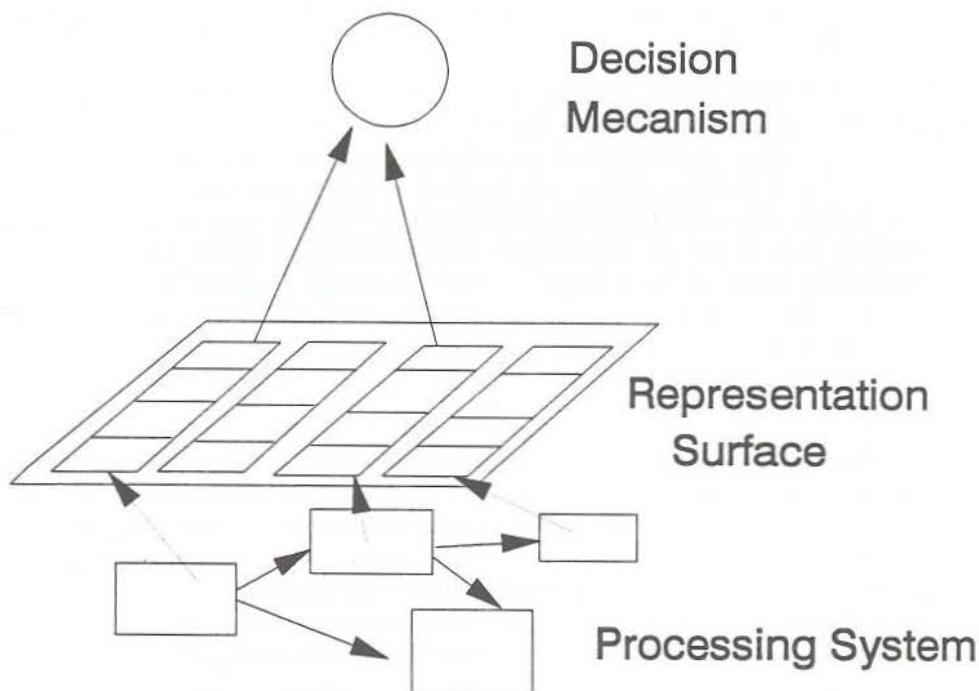
The importance of verification in the determination of subjects' responses is very often overlooked. For instance, Rosch and her colleagues (Rosch, 1973; Rosch & Mervis, 1975) had argued that concepts are represented in terms of prototypes. Indeed, latencies are shorter when subjects are asked to say whether a canary is a bird than to say whether an ostrich is a bird. Likewise, it takes less time to acknowledge that a car rather than a Zeppelin is a vehicle. From such observations, Rosch proposed that concepts are represented in terms of prototypical members from which less typical ones can be retrieved. But, this is not so. Armstrong, Gleitman, and Gleitman (1983) showed that even for concepts that can only be represented in terms of their definition, subjects tend to respond faster to some items than to others. Hence, if one asks subjects to state whether 7 is an odd number they will respond yes much faster than they would if they were asked the question about, say, number 87 though, of course, by definition both are equally odd. The reasons, to make the story short, have nothing to do with the way in which concepts are being represented but rather with the manner in which it is possible to verify whether a token is or is not part of the concept's domain. If this is the case with concepts, the situation is likely to be similar with speech segments.

AN INTERPRETATIVE FRAMEWORK FOR STUDIES OF SPEECH UNITS

To understand the meaning of latency differences in segment monitoring tasks, and for that matter, in many other tasks, one has to have a model that links the processing of the signal, its representation and the constraints under which subjects operate, that is, in particular the verifications that are necessary for a response to be triggered. Even if subjects' responses reflect bottom-up information transduction, it is the cognitive system that makes the decision to respond or not (see Forster, 1979, 1985). Having acknowledged this much, we must also admit that when working with speech signals, the linguistic background of subjects cannot be neglected. Psycholinguists have proposed that different languages rely on representations which are optimally suited to each language (Cutler, Mehler, Norris, & Segui, 1986). In our view, speech

signals are represented in language specific codes, and only these are probably in direct contact with the central verification device. A sketch like the one in figure 1 illustrates what we have in mind.

Figure 1



In our framework, we acknowledge a fundamental distinction between three components: the processing system, the representation surface, and the decision mechanism.

The *processing system* consists of specialized devices that transform information from one representation format or code to another. Examples of these processing subsystems include mapping between retinotopic and object centred representations, between orthographic form and phonemic form, or between word form and word meaning. We assume that the processing system is completely data driven and is opaque for subjects who have no control on the time course of information transduction.

Some of the codes that are involved in these transduction stages have a special status and may appear in the *representation surface*. For each of these codes, a temporary buffer or short term memory store exists that holds traces of preceding events. It is our contention that only codes that appear in the representation surface can be memorised and explicitly manipulated in metalinguistic games or tasks. Examples of such codes include a pre-lexical code like the one used to represent non-words or words which one has never heard before (presumably, the pre-lexical code makes it possible for the child to learn language in the first place); the lexical code, where all the words one knows are represented; and the alphabetic code which provides a written code for incoming stimuli.

Notice that some of the representations and some of the processing devices are entirely cultural while others are not. Unless one has learned to read and write, orthographic representation (or whatever the writing code may be) will be unavailable. Likewise, unless one has had at least a few years of experience with language, the lexical representation will be void or very meagre. Within this sketch, one of the decisive issues is to establish at which level does the diachronically acquired knowledge of one's language begin to play a role in processing, hence in the responses that subjects make in psycholinguistic tests like the ones we use. A more specific question deals with the potential influence of the acquisition of the orthographic code on the other codes and processes.

Last but not least, our framework mentions a third component, which can be called the *decision mechanism*. This component is responsible for polling the information from the different representation surfaces before inhibiting or triggering a response. For each of the codes, we postulate that the representation surface acts as a memory buffer that keeps track of the different events at that level of description. The same event is represented at each one of the available buffers. The cognitive system initiates matching or searching operations each time a stimulus arises and these operations are performed in parallel across all the relevant buffers. A typical experimental protocol consists of the following specification: "If the model in the buffers matches a perceptual event then initiate a response, else freeze". A similar specification can be provided for a classification task. The decision to respond, in our view, involves matching stimulus and target in parallel across all codes and then weight the evidence in order to arrive at a single decision (Forster, 1979).

Before going further, we would like to emphasize that in our model, the only free parameters that can change from one task or experimental situation to the next are the following: the weights that are given to each code and the decision parameters. The decision parameters can be viewed classically as response criterion and bias. The weights reflect the fact that in some tasks, subjects may be biased to rely more heavily on one code or another. In this paper, we will consider that those are the only parameters that can affect the decision mechanism. How these parameters are set by the subjects is not

within the scope of this presentation. Rather, our aim is to show that a mechanical model as this one can serve as a unitary framework to conceptualize on-line studies of speech and studies of awareness.

Within this framework, one can make predictions as to the behaviour of subjects when an inconsistent outcome is given by the different codes. If for instance one code signals a discrepancy between the target and the stimulus, whereas all the other codes signal a perfect match, there will be interference at the decision level that will be reflected as an increase in reaction time. The fact that the speech signal unfolds in time, however, adds another aspect that must be taken into account: It is very likely that all codes do not become active at the same time. So, in addition to patterns of interference between codes, our framework can capture the time dependent or horse race nature of many results in speech perception (see Fox, 1984; Miller & Dexter, 1988; Dell & Newman, 1980; Dupoux, 1989, etc.).

Above, we have presented a very general version of the unitary framework. In the rest of the paper we will choose a particular version, a fairly primitive one, with the purpose of illustrating how it might help to understand a variety of behaviours centred around *chunk detection tasks*. This specific version of the model was developed with a rather narrow purpose, namely, to explore the ways in which literacy can have an effect on chronometric studies. For instance, since we assume that in these kinds of studies, lexical representations do not play a major role (it does not help you to know the meaning of the word *dog* when the task is to decide whether or not it starts with the phoneme *d*), we will neglect them hereafter. Our model assumes a syllabic prelexical code (Mehler, 1981; Mehler, Dupoux, & Segui, 1990) and an orthographic code. However, there are other codes that surely play a role, namely, an articulatory code, a representation of the stress pattern, and so forth. However, we shall not say anything of substance about them in this paper. What is new in our proposal is that we assume that not only is an orthographic code available when listening to speech (see Schneider, Healy, Ericsson & Bourne, 1989), but that this code is available on-line so that it can influence speeded responses to speech stimuli¹.

RESULTS IN SPEECH CHUNK DETECTION

Mehler, Segui, and Frauenfelder (1981) asked subjects to detect segment sequences that occur at the onset of words. For instance, they asked subjects to detect the sequence *pal*² in words like *pa.lace* or *pal.mier*. The main result from this study is that subjects *can* do the task (the error rate is very low), yet they are faster when the target sequence matches the first syllable of the word

¹In the following, we use "orthographic code" to refer to a code which is unavailable to illiterate subjects and which is brought about by a specific training in some orthography. We remain neutral about the "visual" status of such a representation.

²The dot indicates a syllable boundary in French.

than when it does not. In other words, syllabic structure interferes with the detection of a sequence of phonemes. Similar experiments were later carried out in a variety of languages with contrasting results. Our aim in the following sections is not to account for these language specific variations, but rather to concentrate on the French results and try to illustrate how we can account for it within our general framework. In the original study, these results were taken to imply that the syllabic structure of the word being heard is computed on-line and used to elaborate a response. Our current view about the above experiment rests upon the parallel activation of the three representations: lexical, syllabic and orthographic. Given the experimental design, we have every reason to believe that lexical representation does not play a role in subjects responses.

Table 1

TARGET			WORD			MATCH	
Stimulus	Syllabic code	Orthographic code	Stimulus	Syllabic code	Orthographic code	Syll. code	Orth. code
pa	{pa}	{p.a}	palace	{pa.lace}	{p.a.l...}	Y	Y
			palmier	{pal.mier}	{p.a.l...}	N	Y
pal	{pal}	{p.a.l}	palace	{pa.lace}	{p.a.l...}	N	Y
			palmier	{pal.mier}	{p.a.l...}	Y	Y

This leaves us with only two codes: the syllabic and the orthographic (see Table 1). Let us examine the prediction one would make if subjects merely used the syllabic code to perform the task. Of course, it is easy to explain why subjects respond when the target matches the first syllable of the word. However, it is much harder to explain why subjects respond to non-matching cases. One might propose some generic argument suggesting that since there is no complete acoustic mismatch between the syllable /pa/ and the syllable /pal/, the decision mechanism is deceived by such a partial match. Alternatively, if one assumes that subjects use only the orthographic code, it is very hard to understand why there should be a reaction time difference between matching vs non matching cases, though one might propose that the orthographic code includes some information about syllabic structure (see Taft, 1979; Taft & Forster, 1976; Prinzmetal, Treiman, & Rho, 1986; among others).

The richer alternative which we will further explore here proposes that both syllabic and orthographic codes are activated and matched in parallel. The subjects' response is primarily driven by the orthographic code, but we can see that mismatches at the syllabic level will produce interference. In this framework, the basic result can be stated in the following way: Even though subjects can in principle do the chunk detection task on the basis of the orthographic code alone, they cannot suppress the syllabic code which

interferes with their response. Why is this so? The reason may be that the syllabic code is a primary unit for speech perception, and that the orthographic code is only derived later in the course of learning to read. Notice that in what we said so far, nothing hinges precisely on there being an orthographic code per se. However, there could very well be a *phonemic* code. What is important is that such a code represents speech sounds as a purely linear sequence of elements. The reason for referring to it as an orthographic code will become clearer when we examine the performance of other subjects populations.

EFFECTS OF LITERACY

Literacy has a very strong impact on subjects' abilities to perform segment monitoring tasks (Morais et al., 1979). Illiterates find it very difficult if not outright impossible to detect isolated segments; in contrast they perform better with larger parts of words such as syllables. In our model, this difficulty can be accounted for in a straightforward way. Since illiterates have no orthographic code, the only way they can perform metalinguistic tasks is through the use of the sole syllabic representation. But it is easy to see that within a syllabic representation it is very difficult to manipulate or to specify individual segments as targets.

The relevant finding here is that Portuguese illiterates are sensitive to the syllabic structure of words (Morais, Content, Cary, Mehler, and Segui, 1989): They responded significantly more often when the target matched the first syllable of the word than when it did not match. In other words they acted as if they were unaware of the fact that the segment /pa/ is part of the word *palmier* or the segment /pal/ is in the word *palace*. This is exactly the result we would expect (see Table 2) if subjects lack or do not have access to the orthographic code.

Table 2

TARGET		WORD		MATCH
Stimulus	Syllabic code	Stimulus	Syllabic code	Syll. code
pa	{pa}	palace	{pa.lace}	Y
		palmier	{pal.mier}	N
pal	{pal}	palace	{pa.lace}	N
		palmier	{pal.mier}	Y

Of course, in our model, subjects would simply NOT respond to mismatches. However, Morais et al. (1989) observed that illiterates did respond to mismatches although significantly less often than to matches. If this is so, this would mean that illiterates have (limited) access to a sequential

code. Still, to settle the point, we would need to assess the chance level of illiterates—their response rate for similar but different syllables (e.g., *pa* in balance). In the absence of such controls, it is difficult to know whether they have limited access to a sequential code, or no access at all.

In brief, we posit that illiterates rely more heavily on the syllabic code than on a sequential code which is either unavailable (an orthographic code) or very encapsulated (a phonemic code). This simple assumption accounts for two phenomena at once: *a*) illiterates can only perform exact structural matches between target and word and *b*) they cannot do single phoneme detection.

MORE ON ORTHOGRAPHY: THE CASE OF JAPANESE

We turn now to another study using yet another subjects population: Japanese speakers. The Japanese case is interesting because literate speakers do not use an alphabetic code. Instead Japanese speakers use two codes, the Kana code and the Kanji code. The Kanji code is similar to the Chinese system in that it represents both meaning and phonological information in pictographic symbols. The Kana is a phonological code that represents each Japanese *mora* as a separate symbol. Phonologically, a mora is a subsyllabic unit³. For instance, in Japanese, a word like *tanshi* is disyllabic /*tan.shi*/, but trimoraic /*(ta)(n)(shi)*/, where the consonant (*n*) in coda position counts as a single mora and is represented as a single character. In contrast a word like *tanishi* is both trisyllabic and trimoraic /*(ta).(ni).(shi)*/.

What happens then in monitoring situations? Presumably, the Kanji symbols cannot come into play in these experimental settings, for the same reasons that lexical information cannot come into play in the above mentioned experiments (the kanji representation of the target is irrelevant for the task and the kanji representation of the word being heard is not yet available by the time subjects initiate their responses). However, the kana code could come into play as we will see below.

Otake, Hatano, Cutler, and Mehler (in press) had Japanese subjects detect *ta* and *tan* in words like *tanshi* and *tanishi*. They found a pattern of results very different from the French or the Portuguese. One striking result is that subjects find it very difficult to detect *tan* in *tanishi*. Indeed, in 64% of the cases, they simply did not respond. Such a result might be expected if subjects rely on a mora representation. In words like *tanshi*, reaction times to /*ta*/ targets were faster than to /*tan*/ target, again a result to be expected if subjects used a mora representation (one mora to match is faster than two). How can we account for these results in terms of our model?

One possibility might be to replace our previous syllabic code with a mora code. With this hypothesis, Japanese and French subjects differ in the way

³ In Japanese, there are exactly 105 possible morae and they fall into five categories: CV, CCV, V, a nasal consonant when it appears in a coda position, or a geminate consonant.

they encode speech sounds. There is however a much simpler hypothesis: Japanese subjects simply use a moraic instead of an alphabetic orthography. The resulting model is illustrated in Table 3.

One sees here that the missing data point corresponds to cases where there is a mismatch both in terms of syllable and mora (*tan* in *tanishi*). The other comparison involves detection of *ta* versus *tan* in *tanshi*. The prediction in this particular case is less clear, because on the one hand, there is a syllabic mismatch (*ta* in *tan.shi*), while on the other, there are two morae to match instead of one ((*ta*)(*n*) in (*ta*)(*n*)(*shi*)). Which condition is faster depends on the relative weight of these two costs, something that we left unspecified in our model.

Table 3

TARGET			WORD			MATCH	
Stimulus	Syllabic code	Moraic code	Stimulus	Syllabic code	Moraic code	Syll. code	Mor. code
ta	{ta}	{ta}	tanishi	{ta.ni..}	{ta.ni...}	Y	Y
			tanshi	{tan..}	{ta.n...}	N	Y
tan	{tan}	{ta.n}	tanishi	{ta.ni..}	{ta.ni...}	N	Y.N
			tanshi	{tan..}	{ta.n...}	Y	Y.Y

Still, one sees that a fair amount of mileage can be obtained from simply saying that the difference between the French case and the Japanese case is a difference in the orthographic system used to represent speech sounds. Moreover, the above account makes a number of strong predictions.

Our model suggests that Japanese subjects should have a great deal of difficulty in performing initial consonant detection, since initial consonants never correspond to a mora⁴. In contrast, they will detect the same phonemes when they appear in coda position (the *n* in *honda*) because, there, they count as a single mora⁵.

In our account, notice that we still postulate that the Japanese are subject to the classical syllabic interference. We cannot see this interference in the data, however, because it is masked by the massive effect of orthography. One straightforward prediction would be that Japanese illiterates (or preliterate

⁴ Morais et al. (personal communication) indicated that Japanese literates perform metalinguistic games such as inverting the phonemes in words like *uke*. However, they seem to do so by using the kana matrix and performing a complex sequence of operations on it. It is thus unlikely that such a strategy could be used to perform on-line segment monitoring tasks.

⁵ In fact, the only consonants allowed freely in coda position in Japanese are the nasals *n*, *m* and *gn*. There are also geminate stops like in *katta*, but the *t* does not appear as a mora in the orthography. It appears implicitly with a symbol for gemination.

kids) will show a syllabic effect and no mora effect. Another prediction is that alphabetically literate Japanese subjects (who have learned the romanje character set) will lose the mora effect (or show it to a lesser extent) while still showing syllabic effects. All these predictions need to be tested in great detail. In the meantime, it is obvious that the characteristics of the orthographic code being used cannot be neglected when it comes to analyse chronometric studies. Further, our model allows us to classify the different populations on a scale that starts at illiterates who have no orthographic code at all and hence rely exclusively on the syllabic code, then Japanese subjects who have a code that represents subparts of the syllable, and finally subjects who master an alphabetic code that represents individual phonemes. One could ask whether it would be possible to go further than this and find subjects who rely exclusively on a phonemic code. The Spanish case might be one step towards an answer to this question.

THE CASE OF SPANISH: ORTHOGRAPHIC TRANSPARENCY

Sebastian, Dupoux, Segui, and Mehler (1992) reported that when Spanish subjects monitored for speech chunks they failed to show syllabic effects. It was only when subjects were artificially slowed down that a significant interaction between target type and word type emerged. The fact that a syllabic effect was observed with Spanish subjects—albeit under special circumstances—shows that they have a syllabic code, just as the French subjects do. But still there does not seem to be a difference between the two populations concerning what counts as the typical response pattern in chunk detection. For the French it is syllabic, while for the Spanish, it is not. The authors interpreted the difference between French and Spanish in terms of *acoustic transparency*. French has 14 different vowels whereas Spanish has only five. In order to detect a CV or a CVC segment, French subjects presumably need much more acoustic information, hence rely on a large speech chunk (the syllable) in order to extract segments. In contrast, Spanish subjects rely on very local cues and short-circuit the syllabic code.

But there is also another important difference between Spanish and French. Spanish has a very transparent orthography. That is, for a given phoneme there is usually a single grapheme and vice versa. French in contrast is much less regular: The sound /o/ can be written *o*, *au*, *eau*, *ot*, *aud*, and so on. Many letters are not pronounced, and when they are pronounced, the phonemic realisation depends heavily on the context (the letter *c* is /s/ in *cinema*, /ʃ/ in *cheval*, /k/ in *christ*, etc.).

Could the difference between Spanish and French subjects be accounted for in terms of orthographic transparency? In Spanish the transduction between graphemes and phonemes is much more regular and hence faster than in French. Hence, in the chunk detection task, Spanish subjects probably rely more heavily on the graphemic code which becomes available sooner and is more reliable than it is for the French subjects. But if Spanish subjects rely

more heavily on their orthographic code, there should be less interference effect from the syllabic code.

In a nutshell the more phonemically aware you are, the more you focus on the sequential level alone, and the less syllabic effects there should be. Phonemic awareness depends on the writing system that has been learned. The prediction thus becomes: The more phonetically transparent your orthography is, the fewer syllabic effects you should find. Indeed, in Spanish, a very transparent language, syllabic effects are difficult to uncover while, in French, which is less transparent, syllabic effects remain relatively easy to establish.

However, what still remains unclear is why speed of response in the monitoring task should have an effect as to whether the syllabic level can be ignored or not. We suggest that there is a race between the two codes. In Spanish the orthographic code becomes available before the syllabic code. Hence subjects can respond on the basis of the orthographic code without any syllabic interference. In contrast, when responses are slowed down the syllabic code becomes available and interferes with the response. In French, the orthographic code is more complex and takes more time than the syllabic code, so French subjects show syllabic effects at any speed. The difference between Spanish and French, we predict, is thus more a matter of degree rather than all or none. It could be that for very transparent French segments or for very very fast responses, syllabic effects start to wash out.

CONCLUDING REMARKS

In this paper, we have explored the ways in which our framework unifies the results of awareness studies with those obtained through chronometric methods. Our framework offers the possibility of viewing phonemic awareness as a continuum, with illiterates at one end and people trained with very transparent phonemic systems on the other. Clearly, within this framework, users of pictographic systems must have a behaviour similar to that of illiterates when it comes to segment detection tasks. We see it as an open question as to whether the human mind can learn to compile devices that yield sub-phonemic codes like allophones or features. Perhaps highly trained professionals (like phonologists or phoneticians) can indeed represent information that would be otherwise buried in the depth of the computational unconscious and, more importantly rely on such codes in on-line judgments.

However, we do not want to claim that any arbitrary code could be compiled and used to perform on-line tasks. Surely, the representations of speech sounds must reflect properties that are computed by the processing system. For instance, in Japanese, the mora is not just an arbitrary defined unit, but rather, is a psychologically real unit, regardless of orthography. For instance, morae are the units over which timing of utterances is computed (Beckman, 1982; Hoequist, 1983). Moreover, it may also be the case that some orthographic systems are more adequate for one language than for

another. For instance, trying to impose a moraic orthography in French or for that matter a syllabic orthography in English might be less than ideal.

Moreover, what we do not propose is that the only difference between speakers is a difference between the orthographic code they entertain. For instance, English speakers use a quite opaque orthographic code (at least as opaque as the French one), yet they perform on the chunk detection experiment quite differently from the French (Cutler et al., 1986; Cutler, Mehler, Norris, & Segui, 1989). In this case, we have to acknowledge that the contrasting phonology of the languages has an impact on the prelexical code that gets used. Thus, our proposal does not deny that the prelexical code (which we have termed syllabic) can be implemented in different ways according to the language's phonology. But we contend that languages differ in their orthographies as well, and that one should not neglect this when interpreting reaction time data.

In our account, we have given a quite prominent role to the orthographic code. We do not have the data to ascertain that the relevant code is really a *visual* representation of the letters or the characters, instead of the relevant phonological level that the orthography captures. This proposal, however bold and brutal, has the virtue of making a number of strong predictions. As we have stated above, there are a number of other codes that have to be considered to extend the scope of our model. It is quite likely that an articulatory code can also play an important role. Cutler, Butterfield, and Williams (1987) suggested that the abstract syllabic shape (CV vs CCV, vs CVC) can be used by subjects as a way to specify the realisation of phonemic targets (see also Pallier, Sebastian, Felguera, Christophe, & Mehler, in preparation). Pitt and Samuel (1990) argued that a representation of the stress pattern of words is being used in on-line detections.

There is still a long way to go before we can unify all the results that have accumulated in the area of speech perception within a coherent model. Yet, our framework provides a new, more refined way to look at chronometric studies and opens up new vistas on the mental representation of speech sounds. It also incorporates studies of awareness with on-line studies of speech perception in a single structure. Metalinguistic games can be used to reveal the codes that are available in the representation surface; these different codes can in turn be taken into account in chronometric studies in order to reveal the nature of the underlying processing system.

ACKNOWLEDGEMENTS

We would like to thank Juan Segui, Anne Cutler, Nuria Sebastian, Christophe Pallier, Anne Christophe, and Stanislas Dehaan for their help, comments, and fruitful discussion. This research was carried out with the help of Centre National d'Etudes en Télécommunications (CNET), Convention 837BD28 00790 9245 LAA/TSS/CMC, a research grant award from the

International Human Frontier Science Program Organization (HFSP) and a grant "Action intégrée de coopération scientifique et technique franco-espagnole" (1990-1991).

REFERENCES

- Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 263-309.
- Beckman, M. E. (1982). Segmental duration and the "mora" in Japanese. *Phonetica*, *39*, 113-135.
- Bertelson, P. (1986). (Ed.). *The onset of literacy*. Cambridge, MA: MIT Press.
- Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language*, *26*, 406-418.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*, 385-400.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilingualism. *Nature*, *320*, 229-230.
- Cutler, A., Norris, D. & van Ooyen, B. (1990). Vowels as phoneme detection targets. *Proceedings of the International Conference on Spoken Language Processing, Vol 1* (pp. 581-584). Kobe, Japan.
- Dell, G. S., & Newman, J. E. (1980). Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 608-623.
- Dupoux, E. (1989). *Identification des mots parlés: détection de phonèmes et unité prélexicale*. PhD dissertation, EHESS, Paris, France.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper & E.C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale, N.J.: Erlbaum.
- Forster, K. I. (1985). Special purpose computation: All is not one. Commentary/Fodor: Modularity of Mind. *Behavioral and Brain Sciences*, *8*, 9-11.
- Fox, R. A. (1984). Effects of lexical status on phonetic categorisation. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 526-540.
- Hoequist, C. E. (1983). Syllable duration in stress-, syllable-, and mora-timed languages. *Phonetica*, *40*, 203-237.
- Liberman, A. M., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Reading and the awareness of phonetic segments. *Journal of Experimental Child Psychology*, *18*, 201-212.
- Luce, R. D. (1986). *Response Times. Their role in inferring elementary mental organization*. Oxford University Press, New York.
- Mehler, J. (1981). The role of syllables in speech processing: Infant & adult data. *Philosophical Transactions of the Royal Society*, *B 295*, 333-352.
- Mehler, J., Dupoux, E., & Segui, J. (1990). Learning constraints on models of speech perception. In J. Altmann (Ed.), *Cognitive Models of Speech Processing*, Cambridge, Mass: MIT Press.

- Mehler, J., Segui, J., & Frauenfelder, U. (1981). The role of the syllable in language acquisition and perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech*, Amsterdam: North Holland.
- Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 3, 369-378.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323-331.
- Morais, J., Content, A., Cary, L., Mehler, J., & Segui, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes*, 4(1), 57-67.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (in press). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*.
- Pallier, C., Sebastian, N., Felguera, T., Christophe, A., & Mehler, J. (in preparation). Attentional allocation within syllabic structure.
- Parasuraman, R. (1986). Vigilance, Monitoring and Search. In K. R. Boff, L. Kaufman & J. P. T. Homas (Eds.), *Handbook of Perception and Human Performance, Vol. II*, New York: Wiley & Sons.
- Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human, Perception, and Performance*, 16, 564-573.
- Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language*, 25, 461-475.
- Pylshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: Bradford.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York, Academic Press.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behaviour*, 9, 295-302.
- Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E. (1989). Letter decision errors in reading, auditory, and memory tasks. *Journal of Memory and Language*, 28, 400-411.
- Sebastian, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, 31, 18-32.
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 5, 546-554.
- Spoehr, K. T. (1981). Word recognition in speech and reading: Toward a single theory of language processing. In P. D. Eimas & J. L. Miller (Eds), *Perspective on the study of Speech*. Hillsdale, NJ: Lawrence Erlbaum.
- Taft, M. (1979). Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behaviour*, 18, 21-39.
- Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behaviour*, 15, 607-620.

- Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, 24, 320-335.
- Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory and Cognition*, 8, 513-520.