PDFlib PLOP: PDF Linearization, Optimization, Protection

Page inserted by evaluation version
www.pdflib.com – sales@pdflib.com

Understanding Compressed Sentences: The Role of Rhythm and Meaning[a]

JACQUES MEHLER,[b, c] NURIA SEBASTIAN,[d]
GERRY ALTMANN,[e] EMMANUEL DUPOUX,[f, c]
ANNE CHRISTOPHE,[c] AND CHRISTOPHE PALLIER[c]

[c]LSCP, CNRS-EHESS
54 Boulevard Raspail
75006 Paris, France

[d]Facultat de Psicologia
Universitat de Barcelona
Campus de Pedralbes
08028 Barcelona, Spain

[e]Department of Experimental Psychology
University of Sussex
Brighton BN1 9OG, England

[f]France Telecom
6, Place D'Alleray
75015 Paris, France

INTRODUCTION

The language user can recognize words uttered by different speakers, even when they vary their intonation and speaking rate. In this paper we focus on the user's ability to recognize word forms regardless of whether they are spoken fast or slow. Indeed, speech rate is highly variable in natural context. For example, a word uttered in isolation may be twice as long as the same word uttered in the middle of a sentence. However, regardless of the mode in which a word is pronounced, the two resulting acoustic signals activate the same lexical representation. That people have this ability suggests that speech must be coded in a time-invariant fashion.

Psychologists are quite familiar with the study of perceptual invariance for the visual domain.[1,2] When it comes to speech, however, the study of perceptual invariance, in particular that of time, has generated less interest, and we are thus still unable to explain how the acoustic/phonetic processors solve this problem. Cognitive scientists who work in the area of speech recognition generally assume that subjects use their lexical knowledge to normalize the signal. Undeniably, lexical processing does intervene at some level, but we believe that normalization of the signal is a necessary part of prelexical processing, that is, it has to take place prior to the intervention of lexical lookup routines.

272

Indeed, automatic speech recognition systems tend to be effective when the lexicon contains a small number of items. However, they are quite poor when the lexicon contains a large number of entries. Thus, pattern matching to a small set of items might be exempt from the usual problems that arise from changes in time and rate. That, however, is not the general case. Mehler, Dupoux, and Segui[3] have argued that the problem of time invariance has to be placed in the context of language acquisition rather than exclusively within the context of speech processing in adults. Indeed, if time invariance can only be attained through pattern matching of forms stored in the lexicon, it becomes difficult, if not outright impossible, to explain how children can acquire the lexicon of their maternal language to start with. To acquire a lexicon, the many different acoustical renderings of a given word must be placed within one and the same category, presumably one that is allowed by the phonology of the language. If we apply this point of view, infants must first use the input to adjust their invariant prelexical representation so that it fits the language in the environment. Once this task has been solved, they can begin to acquire the word forms used in their language. The only alternative proposal we can imagine would roughly claim that infants or young children have perfect and unlimited memories and that, initially, they represent a huge constellation of acoustical forms, indeed, all the speech signals they have ever been exposed to. At first no organization of these acoustic forms occurs. Later, as the child hears an acoustic signal, and according to some metric or another, he links or assimilates it to the most similar entry already in the constellation. Only when the lexical category itself has become established (one wonders how) does the child recognize new versions of the old entry. Of course, for such a mechanism to work one would have to explain how all the acoustic signals that stimulate the child from this massively unstructured memory space come to be grouped into word forms.

Given all we know about neonates and their abilities, cognitive science should get rid of models that are as unrealistic as the one just described. The proposal sketched at the end of the preceding paragraph should thus not be taken too seriously. Indeed, there are studies suggesting that an intermediary representation between the acoustic signal and the lexicon exists that is already functional in infants who have not yet mastered the lexicon.[4,5] This intermediary code is computed after the speech signal is stripped from variations due to speaker identity, rate of speech, illocutionary force, etc.[6] There is little one can say now about the nature of this intermediary code. Possibly, this representation reflects, in part, universal properties of our endowment and in part language-specific properties that arise when a brain like ours is put into contact with a linguistic environment. While all languages rely on segmental as well as prosodic information, some languages may pay closer attention to the distribution of stress, and others to syllable structure. Regardless of how languages reduce the diversity in the signal into a relevant invariant code, it seems ill-advised to deny the existence of such a code. For our purpose what appears essential is the fact that signals are normalized in order to map the speech signal onto an invariant code.

In this paper we present some exploratory studies to establish how time invariance is achieved when listening to sentences that have very different rates of speech. Although we have carried out a number of experiments in different languages, countries, and teams, all make use of the *speech compression technique* rather than of the fast or slow rates of naturally articulated speech. We grant that our studies are limited because we do not yet understand how the time invariance processes that apply for compressed speech can be extrapolated to unaltered speech. However, there is an advantage to the use of compressed speech: namely, in normal circumstances temporal normalization occurs instantaneously and effortlessly, and is there-

fore difficult to study. The use of speech rates much faster than what can be naturally produced may push the system to its limit, and allow us to observe the time course of normalization.

PERCEPTION OF COMPRESSED SENTENCES

The digitization of speech makes it possible to resynthesize the signal with a different duration while leaving most other parameters of the signal unaltered. Thus, it is possible to take a sentence spoken at a normal rate and reproduce it with half the duration leaving the timbre of the voice and the spectral features relatively intact. The speech compression algorithm we used is based on the pitch synchronous overlap and add (PSOLA) technique;[7] roughly speaking, it takes adjacent pitch periods (in the case of voiced portions) and averages them smoothly; for unvoiced segments the algorithm picks out an arbitrary time window to do the averaging. It is enough to say that this method allows one to double the number of words that are pronounced in a given unit of time with only rare instances of any significant decrement in perceptual performance. Speech compression has furnished psycholinguists with a practical tool to inspect speech perception from a novel vantage point.

The identity of speech segments, if guessing is ruled out, must be computed on the basis of acoustical cues. Many are, however, time dependent. How does the perceptual apparatus overcome these huge variations in speaking rates to pull out the appropriate segments? Miller[8,9] has identified two mechanisms: one, that evaluates speaking rate on the basis of local indices (for instance, syllable duration), and the other that determines segment identity on the basis of both the instantaneous acoustic cues that characterize segments and the speaking rate established by the former mechanism. Miller synthesizes a syllable that comprises a vowel preceded by a /b/ or a /w/ consonant just by changing the transition's slopes. When one measures real speech it becomes apparent, however, that these slopes vary with the rate of speech. Conversely, when subjects listen to an ambiguous syllable, either /ba/ or /wa/, the perceived phonetic identity will be influenced by the duration of the whole syllable, and also, to a lesser extent, by the rate of the preceding context. Miller's studies suggest that speech rate is used to determine the segmental identity of the signal. Moreover, Miller and Eimas[10] have shown that the intercategory boundary depends on speech rate in the same way for adults and infants. This last result suggests that the mechanisms for extracting time invariance are rather primitive and low level. We turn to complementary studies that were carried out using the speech compression algorithm, which makes it possible to present continuous speech rather than just isolated syllables.

A team of investigators from different countries and different linguistic communities, namely, Altmann (English), Dupoux, Christophe, and Pallier (French), Sebastian (Spanish and Catalan), Myagishima (Japanese, personal communication), observed an interesting phenomenon that was at the origin of these compression studies. When naive subjects listen to sentences compressed to 50 percent of their initial duration, they find the first sentences difficult to understand, although the ones that follow become easier and often as natural as the uncompressed sentences themselves. This evolution suggests that subjects *adapt* to altered speaking rates. What are the mechanism and the processes that make such an adaptation possible? In this paper we present studies that explore the perception of compressed speech.

TABLE 1. Percentage of Recalled Open-class
Words in Function of Adaptation Condition

Adaptation Rate (%)	French	English
Nothing	61	23
100	61	21
50	86	44
40	89	27

We use different compression rates and adaptation sentences in the same or in a different language than the test sentences.

In order to experimentally investigate adaptation to high compression rates, we conducted two parallel and complementary experiments, one in French and one in English, in Paris and at Sussex University, respectively. We asked subjects to listen to and transcribe one by one, five sentences compressed to 40 percent (of their initial duration). There were two experimental conditions and two control conditions. In the experimental ones, subjects had previously heard a set of ten different compressed sentences at either 40 percent or 50 percent. In the two control groups, subjects either did not receive adaptation sentences or they heard them with their original duration (100 percent). The English and French sentences were translations of each other, and the overall length in number of syllables was matched across languages. Words with similar relative frequencies were selected in both languages. The material was recorded by the same English–French female bilingual who was very fluent in both languages. The subjects' performance was assessed on the basis of the number of items[g] they reported for each test sentence. All groups listened to the same five test sentences so that any differences in performance must be attributed to the conditions leading up to the test sentences.

The results of the French and English experiment are presented in TABLE 1. First and foremost, notice that the performance of the control groups is very different in the two languages. The two French control groups transcribed correctly over 60 percent of the open-class words, while the two English control groups transcribed roughly 20 percent of the open-class words. It is unclear why performance by these control groups differs so dramatically. We cannot affirm that the testing conditions and the ambient noise were identical in the two laboratories. Neither can we maintain that the bilingual speaker was entirely comparable in the two languages; indeed, the original rate of the English sentences was slightly faster than that of the French sentences. It is also possible that the languages in question and the materials had contrasting effects. At any rate, we believe that the two experiments should be considered separately and that a direct comparison of the two should only be subject to extreme caution.

The performance of the French subjects was alike whether they heard the test sentences without any prior context or after hearing ten uncompressed sentences. However, their performance improved significantly after they had heard compressed sentences, regardless of whether these were compressed to 40 percent, the same rate as the test sentences, or to 50 percent. This result corroborates our subjective

---

[g]Three different scores were considered: the number of words, the number of open-class words, and the number of syllables. No significant effects emerge by scoring one way rather than another.

impression that the ability to perceive compressed speech improves by shear exposure to high rates of speech. The fact that the absence of context yielded similar performances to those obtained with the context of natural sentences is interesting. It suggests that the improvement we measured in the other two conditions is not just a general improvement in doing this kind of a task, such as improving one's transcription skills or getting better at recovering materials stored in immediate memory. Instead, the improvement has to come from more perceptual levels.

The results of the English experiment parallel the results of the French study with the exception of one condition. Indeed, although there was an improvement for the group that got the sentences compressed at 50 percent, there was no sign of improvement for the group that listened to the sentences at 40 percent, namely, the same rate as the test sentences. These contrasting results are, at first glance, rather anti-intuitive. Indeed, why should subjects' performance only improve when they listen to compressed sentences, but at a compression rate that differs from that in the test sentences and not otherwise? One hypothesis that comes to mind is that subjects will adapt, if and only if they can "handle" the context sentences. Indeed, the overall performance on the test sentences hinges, up to a certain point, on how the compressed sentences are initially perceived. If subjects can pay attention to the context sentences and understand many of the words in them, adaptation arises. Presumably when the context sentences were compressed to 50 percent subjects can "handle" the stimuli like normal speech and their performance improves. This, apparently, was not so when the context sentences were compressed to 40 percent. Other studies will tell whether this interpretation is satisfactory or not.

Sebastian carried out experiments in Spanish and Catalan using a similar design, but reducing the number of groups to one experimental (adaptation at 36 percent) and one control (no adaptation sentences). In her case, the test sentences were compressed at 36 percent of their original duration. She found a significantly improved performance for subjects who were adapted with sentences compressed to the same rate as the test sentences (see TABLE 2). Moreover, despite the fact that the test sentences were compressed to nearly a third of their original duration, the performance of the control group is better than 50 percent. As we suggested earlier, with compression rates that allow the controls a performance close to 50 percent adaptation, the presentation of compressed sentences usually improves perception.

So far, the experiments reported before suggest that the adaptation to compressed speech, which is easy to observe informally, is also rather easy to corroborate in the laboratory. Out of four different experiments, we failed to find improvement in intelligibility of compressed speech in only one experiment, the one in which the subjects' initial performance was very low. What we need to address next is the kind of mechanism that subjects use in order to improve their performance. How long does it take for adaptation to set in? What aspect of the adapting stimuli is necessary to obtain improved perception? Dupoux and Green[11] have gathered some data that specifically explores these issues.

TABLE 2. Percentage of Recalled Open-class Words in Function of Adaptation Condition

Adaptation Rate (%)	Spanish	Catalan
Nothing	67	62
36	92	79

Dupoux and Green[11] carried out an experiment to determine the time course of adaptation. To this end, they presented the same five sentences in different serial positions in lists of 20 sentences. This allowed them to examine the performance on a given sentence when it was preceded by fifteen, ten, five, or no compressed sentences. Their study discloses that the more compressed sentences one hears the better one becomes at reporting a given sentence. This slow growing improvement requires ten or more sentences to reach asymptote.

The preceding results mesh badly with the intuitive observations outside of the laboratory. Indeed, naive subjects feel that one or two sentences are sufficient for them to tune into high rates of compression. Dupoux and Green show that this phenomenological experience is unfounded and that adaptation is, in all likelihood, a slow process by which subjects learn to adjust their perceptual processes to the new rates. Such a slow pace could reflect a rather central learning strategy (subjects learn to guess more efficiently). But a slow process is not necessarily a postperceptual or strategic process. An analogy that one could use at this point is a situation where subjects wear distorting optical prisms. Indeed, after a more or less prolonged time of complete clumsiness, subjects who wear these tend to adapt, in the sense that they become able to reach out and grasp objects in their surroundings without making errors they initially could not avoid making. If the prisms are removed, subjects make errors in the reverse direction to what they did when the prisms were first worm. This process may take days or even weeks to reach its peak, yet it is doubtlessly a low-level process, not a central nor a strategic one.

The prism analogy has of course its limitation since when a subject has adapted to compressed speech it is not the case that he or she cannot process normal uncompressed speech anymore. Dupoux and Green have gathered some data showing that if subjects are presented with normal speech after having adapted to compressed speech, they do not go back to their original unadapted performance level when they are again presented with compressed speech. Indeed, subjects tend to keep the same performance level they had before switching to normal speech. So maybe the more correct analogy would be the situation that arises when one is faced for the first time with a very heavy foreign accent. As with the prism, adaptation may take some time, but unlike with the prism, one may retain understanding for this particular signal through an extended period of time. Incidentally, subjects who adapt to very fast rates tend to report that when they go back to normal rates of speech they hear it as being far too slow. Attempts to establish an objective measure of these subjective reports have been so far unsuccessful. Obviously, more work is required before the mechanisms of adaptation can be fully understood.

Be that as it may, we can ask a slightly different question: How similar do the relevant parameters in the induction and test sentences have to be in order for adaptation to arise? One way to study this situation is to explore whether the meaning and the rhythm of the induction sentences have to be processed and represented for adaptation to arise.

Recent studies have suggested that prelexical representations are language specific. In particular, Cutler et al.,[12,13] Sebastian et al.,[14] and Otake et al.[15] have claimed that speakers of French, English, Spanish, Catalan, and Japanese use language-specific processing routines. Indeed, Cutler et al. have shown that native speakers of French are faster when they have to respond to the first syllable in a word (PA in PAlace or PAL in PALmier) as compared to cases in which the target is either one segment longer or one segment shorter than the first syllable. Speakers of English, in contrast, show no such effect regardless of whether they have to listen to English or French stimuli. Similar differences have been found with speakers of

Japanese who make mora-based responses that the English or French speakers do not. The speakers of Spanish use either the syllable or a demisyllable, depending on the speed of their response.

Different languages use different timing units to represent the speech signal. If so, could the commonality of timing units be one of the factors that determines whether there is adaptation or not? For adaptation to arise, must the timing units in the induction and test sentences resemble one another? Would speakers of French find a Spanish context more useful for adapting to fast rates of speech than English? Would speakers of Spanish benefit from Catalan because both are syllable-timed languages? Obviously, to answer such questions, we have to control for how familiar subjects are with the induction language. Indeed, whether subjects understand or not the context language may have a vast effect on performance. Suppose subjects listened to Jabberwocky-like sentences, or, for that matter, to a text by Derrida. Would they show signs of adapting to faster rates? If they did, one would have the demonstration that understanding the sentences is unnecessary for adaptation to arise. If they do not, we would have to conclude that the process of understanding is an essential component for adaptation to take place.

CROSS-LINGUISTIC COMPRESSION STUDIES

As we argued earlier, adapting to compressed speech is a phenomenon that takes time to reach asymptote and seems, at least in part, located on the level of prelexical representations. To establish whether this is the case, one needs to evaluate the improvement in one language after subjects listen to induction sentences in either the same or in a different language. Subjects may benefit from prior stimulation, regardless of the induction language. If so, one might conclude that speech normalization to very fast speech uses universal representations. Alternatively, subjects may be able to benefit from the induction sentences only if they are related to the test sentences. If this were the case one might argue that adaptation rests on processing structures that are language specific. If so, we might be able to isolate natural language families in terms of the prelexical units they employ. Maybe all the languages that use the same prelexical units might promote adaptation, regardless of whether subjects understand the language or not.

Below we present studies on two kinds of populations, monolingual subjects of different linguistic communities and bilingual subjects.

Monolingual Subjects

The studies reported in the first part of this section were carried out using French and English monolingual subjects who were tested in Paris and at Sussex University. The experimental design is quite similar to the one presented earlier, namely, subjects of the experimental groups are tested on five compressed sentences after having been habituated on ten induction sentences. Subjects have to write down all they can recall about the test sentences, while they simply have to listen attentively to the induction sentences. All subjects were students between 20 and 25 years old.

The results for the French subjects were quite unequivocal. The subjects who listened to English compressed induction sentences performed like the control group who heard the compressed test sentences without any induction context. Likewise, subjects who heard the induction sentences at a natural rate performed as poorly as

the controls, who, as we know, only heard the test sentences. These results suggest that French subjects do not improve their ability to perceive compressed French sentences by being exposed to compressed English sentences.

Similarly, English subjects tested in the converse situation did not improve their performance when presented with compressed French sentences. Unfortunately, the interpretation of these results is clouded by the fact that as we mentioned before, English subjects did not improve their performances in the condition where they were presented with compressed English. To understand what is really going on, another study with controls who have higher base performance rates would be needed in order to ascertain that the obtained pattern of results is truly genuine.

The picture can be extended, however, by considering a similar experiment conducted by Sebastian and described as follows. She showed that Spanish subjects, adapted to compressed Catalan sentences, perform significantly better on compressed Spanish sentences than the control group, which received the test sentences without having been subjected to the compressed context. Last but not least, Miyagishima (personal communication) showed that French monolingual subjects do not improve their performance on five compressed French sentences, as compared to a control group, after listening to compressed Japanese. Thus, there is no benefit when the French are habituated to either Japanese or English (if one wants to be lenient and draw this conclusion from the previously reported experiment), but there are benefits when the Spanish subjects are habituated to Catalan.

The tentative results just reported are compatible with, at least, the following two interpretations: (1) *Adaptation takes place at the prelexical level.* We know that French does not rely upon the same prelexical representations as Japanese or English,[9,12] and the failure to observe adaptation when one of these two languages is used as context might be attributed to the fact that the adaptation process or mechanism is determined by the prelexical representations. The Spanish results would then suggest that Spanish and Catalan rely on similar prelexical representations. (2) *Adaptation depends on understanding the words.* Spanish and Catalan have many words that derive from the same roots; at any rate, many more than French and English or Japanese. This last interpretation might seem unlikely, however, since Sebastian reported that Spanish monolinguals were generally unable to identify more than six words out of the ten test sentences.

What interpretation (2) suggests is that understanding the adaptation sentences is essential for habituation to take place. If so, bilingual subjects should benefit from the compressed context sentences even when they are in a different language than the test sentence, providing they understand both languages.

Studies of Compression with Bilingual Subjects

In a series of studies, English–French bilingual subjects were tested with a design that is similar to the one used with monolingual subjects. As in the study with monolingual subjects, we failed to observe any substantive improvement on English test sentences regardless of whether the subjects had been tested in a cross-language condition or not. Subjects, of course, understood both languages perfectly well. This result confirms that regardless of comprehension, habituating to French or English does not help subjects to report sentences in English. Likewise, when subjects have to process French test sentences, adapting to English context sentences does not help at all. Insofar as we can see, these observations could have arisen for one of two reasons. First, when one adapts to a language one uses the phonological representa-

tion of that language to draw generalizations. If the representations of context and test languages do not coincide, there will be no savings when one is habituated to one and tested in the other. Second, for adaptation to arise, one has to be able to cope with the context sentences as if they were speech, namely, at rates that put the controls at more or less 50 percent correct. The English sentences, however, were hardly comprehensible even by monolingual native speakers of English. Thus, we cannot choose between these two alternatives until a new study becomes available.

Spanish and Catalan, as opposed to English and French, have quite similar phonologies. Indeed, as we pointed out earlier, both languages are, at a first approximation, syllable-timed languages; moreover we know from a previous study that the control groups perform rather well on both types of stimuli.

The Spanish–Catalan bilingual study that was run by Sebastian provides unambiguous evidence of improved performance on test sentences regardless of whether the bilingual subjects were presented with habituation sentences in one language or the other. This result suggests that both Spanish and Catalan rely on rather similar phonological representations, at any rate, at the level of description that is relevant to understand performance in this task. Alternatively, it could be that understanding language is an essential part of the process of rate adaptation. It should be recalled, however, that the monolingual Spanish subjects who listen to Catalan induction sentences also improve their performance on the test sentences. These subjects did not have any understanding of spoken Catalan. Even though the roots of many of the open-class items might have been similar, the subjects were unable, as was mentioned before, to understand these roots when they received a Catalan pronunciation. Yet, the monolingual adaptation data look very similar to the bilingual data presented earlier, suggesting that full understanding may not be a prerequisite for adaptation to compressed speech. Still, as we stated before, several essential studies are necessary before we have a better understanding of the processes that underlie the adaptation to fast rates.

CONCLUSION

The series of studies previously presented all have a common aim, namely, to establish how the human brain processes the speech signal and frees itself of potential misinterpretations, incomprehension, and other major obstacles. These studies represent a small set of the future studies that will be necessary to understand how the human brain computes time invariance from the speech signal. Even to draw preliminary conclusions we will need more data than are currently available.

Yet these studies have, by and large, demonstrated that the language user adapts to exceptionally high compression rates. This result concurs with the incidental observations mostly carried out in the open. It is difficult to state that listening to compressed speech engages exactly the same kinds of processes that one encounters when listening to speakers who use very rapid rates. However, barring such a reservation, it still is the case that subjects report hearing a sentence that is spoken with recognizable timbre, but only very, very rapidly. Indeed, subjects do not report hearing distorted or artificial sounding speech for large ranges of compression. Thus, one can speculate that the language user has little if any difficulty in processing speech, even when its temporal properties are entirely novel. How the brain manages to make the necessary computations is what the preceding experiments have tried to clarify.

From the results reported, it appears that subjects slowly learn to cope with

compressed speech at relatively high rates. It was shown that to become asymptotic subjects may require up to ten sentences of practice. Furthermore, it appears that the induction sentences need not be comprehensible to subjects in order for the adaptation or learning to be effective. If comprehension is not an essential ingredient for learning, what other parameters are essential? We explored whether or not the language phonology of the induction sentences and of the test sentences have to be closely related to observe improved performances. Unfortunately, the results so far available are insufficient to settle the issue. Experiments in progress should clarify the critical case of English versus French. Indeed, if language phonology matters, a syllable-based language like French should not help speakers to adapt to a stress-based language like English. So far we have tested several groups, but the outcome of these studies remains uninterpretable because the control groups in the different languages perform very differently. Conversely we should predict that French and Spanish should adapt each other, since they both are syllable-based languages.

Thus, it may be the case that the technique allows us to define families of languages sharing processing strategies: any pair of languages drawn from the same family should adapt each other, but not languages of disjoint families. If this turns out to be the case, the study of speech compression would become one of the central methods for establishing the representations on which subjects rely when they process speech. Only more studies will uncover whether or not this is the case.

ACKNOWLEDGMENTS

REFERENCES

1. HELMHOLTZ, H. 1921. Schriften zur Erkenntnistheorie. Julius Springer. Berlin.
2. BORING, E. G. 1942. Sensation and Perception in the History of Psychology. Appleton. New York.
3. MEHLER, J., E. DUPOUX & J. SEGUI. 1990. Constraining models of lexical access: The onset of word recognition. *In* Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives, G. Altmann, Ed. MIT Press. Cambridge, Mass.
4. BIJELJAC-BABIC, R., J. BERTONCINI & J. MEHLER. How do four-day-old infants categorize multisyllabic utterances? Dev. Psychol. In press.
5. BERTONICINI, J. & J. MEHLER. 1981. Syllables as units in infant speech perception. Infant Behav. Dev. **4:** 247–260.
6. KUHL, P. 1983. Perception of auditory equivalence classes for speech in early infancy. Infant Behav. Dev. **6:** 263–285.
7. CHARPENTIER, F. J. & M. G. STELLA. 1986. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. Proceedings of the IEEE International Conference ASSP, Tokyo: 2015–2018.
8. MILLER, J. L. 1981. Effects of speaking rate on segmental distinctions. *In* Perspectives on the Study of Speech, P. D. Eimas and J. Miller, Eds. Erlbaum. Hillsdale, N.J.
9. ———. 1985. Rate-dependent processing in speech perception. *In* Progress in the Psychology of Language, A. Ellis, Ed. Erlbaum. London.
10. MILLER, J. L. & P. EIMAS. 1983. Studies in the categorisation of speech by infants. Cognition. **16:** 135–165.
11. DUPOUX, E. & K. GREEN. Adaptation to compressed sentences. In preparation.

12. CUTLER, A., J. MEHLER, D. NORRIS & J. SEGUI. 1983. A language specific comprehension strategy. Nature **304:** 159–160.
13. ———. 1986. The syllable's differing role in the segmentation of French and English. J. Mem. Lang. **25:** 385–400.
14. SEBASTIAN, N., E. DUPOUX, J. SEGUI & J. MEHLER. 1992. Contrasting syllabic effects in Catalan and Spanish. J. Mem. Lang. **31:** 18–32.
15. OTAKE, T., G. HATANO, A. CUTLER & J. MEHLER. Mora or syllable? Speech segmentation in Japanese. Submitted for publication in J. Mem. Lang.