Research Article

Linguistic Constraints on Statistical Computations

The Role of Consonants and Vowels in Continuous Speech Processing

Luca L. Bonatti,[1,2] Marcela Peña,[1,3] Marina Nespor,[4] and Jacques Mehler[1,5]

[1]*International School for Advanced Studies, Trieste, Italy;* [2]*University of Paris VIII at Saint Denis, Saint Denis, France;* [3]*Laboratorio de Neurociencias, Escuela de Psicología, Pontificia Universidad Católica de Chile, Macul, Chile;* [4]*University of Ferrara, Ferrara, Italy; and* [5]*Laboratoire de Science Cognitive et Psycholinguistique, CNRS and EHESS, Paris, France*

ABSTRACT—***Speech is produced mainly in continuous streams containing several words. Listeners can use the transitional probability (TP) between adjacent and non-adjacent syllables to segment "words" from a continuous stream of artificial speech, much as they use TPs to organize a variety of perceptual continua. It is thus possible that a general-purpose statistical device exploits any speech unit to achieve segmentation of speech streams. Alternatively, language may limit what representations are open to statistical investigation according to their specific linguistic role. In this article, we focus on vowels and consonants in continuous speech. We hypothesized that vowels and consonants in words carry different kinds of information, the latter being more tied to word identification and the former to grammar. We thus predicted that in a word identification task involving continuous speech, learners would track TPs among consonants, but not among vowels. Our results show a preferential role for consonants in word identification.***

Adults and infants can perform complex statistical computations to detect the boundaries of words in connected speech. In particular, they have been shown to use statistics to discover nonsense words in continuous streams of artificial speech (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, & Tunick, 1997). However, human statistical abilities span much larger domains than language proper, as statistical calculations can also be used to identify units in sequences of tones, or of visual stimuli (Fiser & Aslin, 2002; Saffran, Johnson, Aslin, & Newport, 1999). It appears that whenever statistical dependencies exist, learners can extract them.

These results still leave open the possibility that specific linguistic factors influence the computation of transitional probabilities (TPs) in speech segmentation. Speech is not only a linear succession of sounds, but also a stimulus that elicits a series of representations, including phonological representations such as syllables, vowels, or consonants, that the brain processes as such from the very early onset of language acquisition (Dehaene-Lambertz & Dehaene, 1994). The specific nature of these representations may constrain the way statistical computations are performed.

One way to investigate this issue is to ask whether TPs can be calculated on elements that, although not phonetically adjacent, are adjacent at some abstract level of representation. For example, in order to account for various morphological and phonological phenomena, it has been proposed that consonants and vowels are represented in separate tiers (Goldsmith, 1976). Accordingly, Newport and Aslin (2004) asked whether learners exposed to a continuous stream of artificial speech consisting of consonant-vowel (CV) syllables could track TPs between consonants and TPs between vowels independently. The results indicated that both kinds of TPs could be used to segment the speech stream. Potentially, therefore, TP computations on both the consonantal and the vocalic tier could help listeners discover words in fluent speech.

Newport and Aslin's (2004) work exemplifies well one way to understand the relation between language and statistical abilities. According to this approach, language essentially provides

Address correspondence to Luca L. Bonatti, SISSA/ISAS, 2-4 Via Beirut, 34014 Trieste, Italy; e-mail: lucabonatti@mac.com.

representations for statistical computations. Linguistic input populates the mind with a set of representations, and then statistics can be computed over them. Thus, if speech triggers the construction of a vocalic tier and a consonantal tier, then TPs can be computed over the two kinds of representations equally effectively.

There is another way to understand how language and statistical abilities interact, however. That is, language may not only provide representations but also impose constraints that optimize processing given its internal architecture. Such constraints might determine which representations are open to statistical computations. In particular, some elements of speech have a specifically grammatical role, and hence may be processed independently of their local statistical distribution, whereas others better serve the recognition of lexical patterns, and may be subject to distributional analysis on-line.

The functional difference between consonants and vowels offers a test case for this hypothesis. From the point of view of a statistical computational mechanism, consonants and vowels are units as good as any others. However, their role in language is very different. On the basis of several linguistic sources of evidence, Nespor, Peña, and Mehler (2003) proposed that vowel alternation concerns quantity (i.e., duration, pitch, and intensity, which are in different degrees responsible for prosody), whereas consonant alternation concerns quality (as realized, e.g., in different manners or places of articulation).

Several phenomena (e.g., vowel harmony and centralization of unstressed vowels) have the effect of reducing the contrastive power of vowels, impoverishing their role in distinguishing lexical items. Consonants undergo opposite phenomena: Word-internal nonadjacent consonants that are too similar become more distinct in many languages. The Semitic languages are an extreme case that attests to the role of consonants in making lexical distinctions. In these languages, lexical roots are formed exclusively by consonants, whereas vowels are inserted to indicate morphological patterns (McCarthy, 1985). No languages with the opposite pattern are documented; that is, in no known language are lexical roots composed exclusively of vowels and consonants used primarily to provide morphological information. In short, phonological phenomena decrease the potentially distinctive role of vowels and increase that of consonants.

Being the main carriers of prosody, vowels provide cues to syntactic constituency through prominence and lengthening, and in some languages like Turkish, through vowel harmony. In addition, the percentage of vowels contained in a string may provide information about important grammatical properties of language. A high percentage of vowels is typical of languages in which complements precede their heads and agglutinative morphology is used, whereas a low percentage of vowels is typical of languages in which heads precede their complements and there is no agglutination (Shukla, Nespor, & Mehler, 2004). Thus, quantitative information carried by vowels gives important information about syntactic structure, as well as about some aspects of the morphological system. In contrast, the role of consonants in signaling syntax is minor and limited to juncture phenomena that signal constituency (Nespor & Vogel, 1986; Selkirk, 1984).

A categorical distinction between vowels and consonants is also supported by several other lines of evidence. In early language learning, the individuation of vowels and the individuation of consonants as phonological rather than phonetic elements of the infant's native language follow different developmental time courses (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994; Werker & Tees, 1984). Furthermore, studies on brain-damaged patients signal asymmetries in the processing of consonants and vowels in both recognition and production (Caramazza, Chialant, Capasso, & Miceli, 2000).

In short, both theoretical and empirical considerations suggest a functional difference between vowels and consonants. We propose that the categorical distinction between vowels and consonants makes each specialized for a specific function in language processing and acquisition. Vowels give cues mainly about syntax, whereas consonants give cues mainly about the lexicon.

Can this difference affect how distributional information is processed in on-line speech segmentation? If our hypothesis is correct, it is possible that even in an unsegmented, meaningless speech stream, vowels and consonants elicit different computations. As TP computations in an unsegmented speech stream serve mainly to identify potential lexical segments, but are of limited use in discovering underlying grammatical regularities (Peña, Bonatti, Nespor, & Mehler, 2002), we predict that TPs may be calculated on nonadjacent consonants but, in general, cannot be as successfully exploited with nonadjacent vowels.

In this article, we report a series of experiments on adults exposed to continuous streams of artificial languages composed of CV syllables. As in Newport and Aslin's (2004) studies, these streams contained imaginary words delimited by dips in TPs either between nonadjacent consonants or between nonadjacent vowels. The main question of interest was whether participants would be equally able to identify the words regardless of the items over which TPs were defined, or whether the ability to extract TP information would vary according to whether TPs were defined over consonants or vowels.

EXPERIMENT 1

Method

Participants

Fourteen French college students were tested and received $5 for their participation (as did all participants in the experiments reported here).

Stimuli

Two 7-min streams of continuous artificial speech were constructed by pseudorandom concatenation of 12 nonsense trisyllabic items,

TABLE 1
Stimuli Used in Experiment 1

Families and items in continuous speech	Test items	
	Words	Part-words
p_R_g_		
puRagi	puRagi	gibydo
puRegy	puRegy	gymali
poRagy	poRagy	kapoRe
poRegi	poRegi	kemaly
b_d_k_		
biduka	biduka	tobidu
bidoke	bidoke	tupoRa
byduke	byduke	Ragime
bydoka	bydoka	Regybi
m_l_t_		
malitu	malitu	dukapo
malyto	malyto	dokema
melito	melito	litupo
melytu	melytu	lytobi

hereafter referred to as "words" (these items are listed in Table 1). These words were constructed from 18 CV syllables (9 consonants and 6 vowels) and were grouped into three different families. Each family included 4 words, containing the same sequence of consonants. For instance, the family /p_R_g_/ (where _ indicates the occurrence of a vowel) consisted of the words /puRagi, /puRegy/, /poRagy/, and /poRegi/. The words were concatenated with the restriction that the same word or two members of the same family could not be presented twice in succession. In order to keep TPs between adjacent syllables constant, we allowed each specific word to be followed by only four other words, as illustrated in Figure 1. Thus, the TPs between consonants were equal to 1.0 within words and to .5 at word boundaries. The TPs between adjacent and nonadjacent syllables and, more important, between vowels were set to .5 both within and between words. A continuous sound file was generated with the MBROLA text-to-speech software (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996), using French diphones, digitalized in 16-bit mono files at 22050 Hz. To avoid the presence of prosodic cues at word boundaries, we equalized all phonemes by setting their duration at 116 ms and

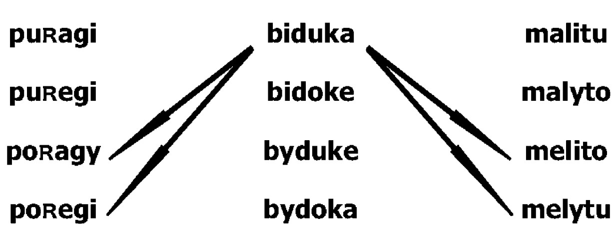

Fig. 1. Example of pseudorandom concatenation of the words in Experiment 1. The arrows indicate the words allowed to occur after the item "biduka."

their pitch at 200 Hz and maintained similar intensity ranges within and between words. No known acoustic cue signaled word boundaries.

For the test phase, 12 trisyllabic items containing the low TPs between consonants were constructed. We call these items "part-words" because they were constructed by joining two syllables of a word to one syllable of the word that immediately preceded or followed it during familiarization. Part-words spanned word boundaries, and hence had lower consonant TPs than words.

Procedure

All participants were tested in a silent room, wearing headphones. They were informed that they would listen to an artificial language containing imaginary meaningless words, and would later be tested about their knowledge of the words of that language.

Before the experiment, in order to ascertain that participants understood the two-alternative forced-choice paradigm, we presented 10 pairs of monosyllables. Participants had to indicate which was the target monosyllable in each pair. Participants who made more than two errors in this pretest phase were excluded from analysis. The experiment was presented on a Pentium-based computer using the experimental software EXPE6 (Pallier, Dupoux, & Jeannin, 1997).

In the familiarization phase, participants heard two randomly presented sequences of artificial continuous speech lasting 7 min each; the two streams were separated by 2 min of silence. In the test phase, participants heard pairs of trisyllabic items, with the members of each pair separated by 500 ms of silence. Each pair contained a word and a part-word from the nonsense language. Each word was tested against three different part-words, for a total of 36 pairs; the serial position of the word within each pair was counterbalanced across subjects. No pair was either preceded or followed by another pair containing one of its members. The list of test items used to create the test pairs is given in Table 1. Participants had to indicate on a keyboard which item in each pair "looked" more like an imaginary word of the language they had listened to. The intertrial interval was 2,000 ms.

Results and Discussion

Participants preferred words over part-words (87.7 $\pm$ 10.6% mean preference for words), $t(13) = 13.23, p < .0001, d = 3.5$ (Fig. 2a). As the words had higher consonant TPs than the part-words did, this result shows that adults can exploit TPs between consonants to segment a continuous speech stream, and replicates Newport and Aslin's (2004) findings with different materials and a different design. The result suggests that the consonantal tier plays a role in the individuation of lexical items in language processing. Next, we tested whether vowels can play a similar role.

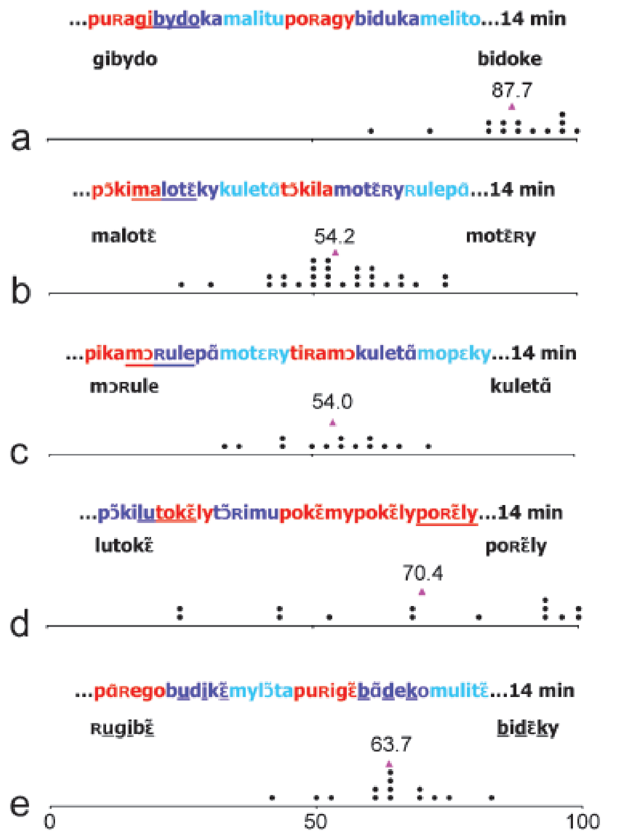

Fig. 2. Samples of the familiarization streams and results from the experiments. In each panel, the first line contains a sample of the 14-min familiarization stream. An example of a test pair is shown above each graph. The dots represent mean preference for words of individual subjects averaged across items; the number above the triangular mark indicates the general mean. Experiment 1 (a) tested participants' ability to segment the stream on the basis of the transitional probabilities (TPs) among consonants. Experiments 2a (b) and 2b (c) tested participants' ability to extract words on the basis of the TPs among vowels. In these three panels, successive words are printed in different colors, words in different colors are from different word families, and underlining identifies examples of part-words used in the test phase. Experiment 2c (d) tested participants' ability to extract words on the basis of TPs among vowels when the number of vowel word families was reduced and consecutive repetitions of families were allowed in the familiarization stream. In this panel, words in different colors represent different families, and the long stretch of red phonemes shows an example of family repetition. Experiment 3 (e) tested participants' recognition of items depending on whether the items maintained the consonantal or the vocalic tiers presented during familiarization. The familiarization stream in this experiment could be segmented on the basis of vowel, consonant, or syllable TPs. Different familiarization items are in different colors. The underlined phonemes are examples of vocalic tiers and consonantal tiers used to construct the test items.

EXPERIMENT 2

Experiment 2a

Method

Twenty-eight French college students participated. The procedure was identical to that in Experiment 1.

TABLE 2

Stimuli Used in Experiment 2a

Families and items in continuous speech	Test items	
	Words	Part-words
_ɔ̃_i_a		
pɔ̃kima	pɔ̃kima	malotɛ̃
pɔ̃ʀila	pɔ̃ʀila	lamopɛ̃
tɔ̃kila	tɔ̃kila	lamotɛ̃
tɔ̃ʀima	tɔ̃ʀima	malopɛ̃
_o_ɛ̃_y		
mopɛ̃ky	mopɛ̃ky	kytɔ̃ʀi
motɛ̃ʀy	motɛ̃ʀy	ʀypɔ̃ki
lopɛ̃ʀy	lopɛ̃ʀy	pɛ̃kyʀu
lotɛ̃ky	lotɛ̃ky	tɛ̃ʀyku
_u_e_ɑ̃		
kumepɑ̃	kumepɑ̃	pɛ̃ʀyku
kuletɑ̃	kuletɑ̃	tɛ̃kyʀu
ʀumetɑ̃	ʀumetɑ̃	mepɑ̃lo
ʀulepɑ̃	ʀulepɑ̃	letɑ̃mo

We constructed the stimuli much as in Experiment 1, but inverted the TP relations between vowels and consonants. The two 7-min continuous streams of artificial speech were created by pseudorandom concatenation of 12 nonsense trisyllabic words (Table 2). As in Experiment 1, words were defined on the basis of TPs between nonadjacent elements of the same category and were constructed out of 18 CV syllables (9 vowels and 6 consonants[1]) grouped into three different families. However, unlike in Experiment 1, each family had four words containing the same sequence of vowels. For instance, the family /_ɔ̃_i_a/ (where _ indicates the occurrence of a consonant) consisted of the words /pɔ̃kima/, /pɔ̃ʀila/, /tɔ̃kila/, and /tɔ̃ʀima/. French speakers discriminate the vocalic differences in these items clearly. Apart from these differences, the material was constructed following the same acoustic and randomization constraints used for Experiment 1. In the continuous speech stream, TPs between vowels were equal to 1.0 within words and to .5 at word boundaries. The TPs between adjacent and nonadjacent syllables and, more important, between consonants were set to .5 both within and between words. For the test, 12 part-words were constructed as in Experiment 1, but the vowel TPs in part-words were lower than the vowel TPs in words (.5–1.0 in part-words, and 1.0 in words).

Results

Despite a sample size twice as big as that of Experiment 1, we found no preference for either words or part-words (54.2 ±

[1]We decided to use CV syllables because switching vowels with consonants within syllables is impossible in continuous speech. VCVCVC sequences are naturally perceived as streams of CV sequences.

11.7% mean preference for words), $t(27) = 1.8$, $p \geq .07$, $d = 0.35$ (Fig. 2b). Common analyses of Experiments 1 and 2 showed a sharp difference between the two means, $t(40) = 8.96$, $p < .0001$, $d = 2.9$.

Experiment 2b

In order to ensure that the failure to extract words in Experiment 2a was not due to idiosyncrasies in our materials, we ran a variant of the experiment using different vowels but maintaining the probability structure of Experiment 2a.

Method

Fourteen French college students participated. Stimuli were prepared following the same constraints as in Experiment 2a, but with different words (Table 3). The procedure was the same as in Experiment 2a.

Results

As shown in Figure 2c, mean preference for words did not differ from chance (54.0 ± 11.3% mean preference for words), $t(13) = 1.3$, $p \geq .21$, $d = 0.35$. The mean was equal to that obtained in Experiment 2a, $t(40) = 0.05$, $p = .9$, $d = 0.01$, but lower than that of Experiment 1, $t(26) = 8.12$, $p < .0001$, $d = 3.07$.

Experiment 2c

Unlike Newport and Aslin (2004, Experiment 3), we found that participants in Experiments 2a and 2b clearly failed to exploit vowel TPs to discover words in continuous speech. Several factors could explain the difference in the results obtained in these studies. One prominent factor is the structural difference between the languages used. In Experiments 2a and 2b, the artificial languages contained three word families with four words each, and in the speech stream, consecutive repetitions of the same word and of the same family were avoided. In Newport and Aslin's Experiment 3, the language contained two families with eight words each, so consecutive repetitions of the same word family necessarily occurred frequently within short time frames. It is possible that such repetitions enhance acoustic or memory cues that can facilitate the identification of word families, quite aside from the probability relationship among vowels. In Experiment 2c, we used a language and materials as similar as possible to those of Experiments 2a and 2b, except that the language contained two families with eight words each and consecutive repetitions of the same family were allowed, as in Newport and Aslin's (2004) Experiment 3.

Method

Fourteen adult French speakers participated. The materials were prepared following the constraints Newport and Aslin (2004) described for their Experiment 3. Two 7-min text streams of artificial continuous speech were constructed by pseudorandom concatenation of 16 nonsense trisyllabic words (Table 4). Words were built out of 12 CV syllables (containing 6 vowels and 6 consonants) and were defined on the basis of the TPs between

TABLE 3
Stimuli Used in Experiment 2b

Families and items in continuous speech	Test items	
	Words	Part-words
_i_a_ɔ		
pikamɔ	pikamɔ	mɔʀule
piʀalɔ	piʀalɔ	lɔkule
tikalɔ	tikalɔ	kamɔʀu
tiʀamɔ	tiʀamɔ	ʀalɔku
_o_ɛ_y		
mopɛky	mopɛky	kykule
motɛʀy	motɛʀy	kyʀume
lopɛʀy	lopɛʀy	pɛkyti
lotɛky	lotɛky	tɛʀypi
_u_e_ã		
kumepã	kumepã	pãlotɛ
kuletã	kuletã	tãmotɛ
ʀumetã	ʀumetã	mepãlo
ʀulepã	ʀulepã	letãmo

TABLE 4
Stimuli Used in Experiment 2c

Families and items in continuous speech	Test items	
	Words	Part-words
_ɔ̃_i_u		
pɔ̃kilu		
pɔ̃kimu		
pɔ̃ʀilu		
pɔ̃ʀimu		
tɔ̃kilu		
tɔ̃kimu		
tɔ̃ʀilu	pɔ̃kilu	kɛ̃mypɔ̃
tɔ̃ʀimu	tɔ̃ʀimu	lutokɛ̃
_o_ɛ̃_y	poʀɛ̃ly	mypɔ̃ʀi
toʀɛ̃my	tokɛ̃ly	ʀiluto
toʀɛ̃ly		
tokɛ̃my		
tokɛ̃ly		
poʀɛ̃ly		
poʀɛ̃my		
pokɛ̃ly		
pokɛ̃my		

Note. In this experiment, four words and four part-words were paired in the test phase, following Newport and Aslin (2004).

their vowels. Two families of words were used, each composed of 8 words with the same sequence of vowels but different consonants. As in Experiments 2a and 2b, two 7-min continuous streams were constructed by controlling TPs between vowels (set to 1.0 within and to .5 between words), TPs between adjacent syllables and consonants (set to .5 within and between words), and immediate word repetitions (never allowed). However, unlike in Experiments 2a and 2b, consecutive repetitions of the same family were allowed, with different members of the same family possibly occurring up to six times consecutively. For the test, 4 of the 16 words were paired exhaustively with 4 part-words to form 16 word/part-word pairs, each presented twice in different orders, for a total of 32 test items (see Table 4).

Results

Participants preferred words over part-words (70.5 ± 27.8% mean preference for words), $t(13) = 2.7$, $p \leq .016$, $d = 0.73$ (Fig. 2d). Mean preference for words was higher than in Experiments 2a and 2b, $t(40) = 2.7$, $p \leq .01$, $d = 0.88$, and $t(26) = 2.1$, $p \leq .049$, $d = 0.78$, respectively, but lower than in Experiment 1, $t(26) = -2.16$, $p \leq .041$, $d = 0.81$.

Discussion

Experiments 1, 2a, and 2b document a striking difference in how language learners can exploit the same statistical information depending on whether it is carried by consonants or by vowels. Although adults are able to break a continuous speech stream into its component words when relying on consonants, they are apparently unable to do so when relying on vowels.

As Experiment 2c suggests, under highly redundant conditions of presentation, vowels also may become perceptually salient and facilitate word recognition, although the effect is less than that of consonants in nonredundant conditions. However, listeners' success in these redundant conditions could arise because consecutive repetitions enhance memory for acoustically similar patterns, rather than because listeners perform statistical calculations over segments of the stream. When a slightly higher degree of family variation is introduced, as in Experiments 2a and 2b, vowel TPs cease to be useful for word segmentation, but consonant TPs are still highly effective.

Why are consonants more transparent to TP computations than vowels? We suggest that the difference may be due to the role of consonants in word identification. If consonants are sufficient to identify words' lexical roots, then it makes functional sense for the language-learning device to use consonants, but not vowels, to segment streams of speech. This hypothesis implies that when learners compute TPs over consonants, they will tend to extract, not the actual sequence of syllables that compose the words encountered in the stream, but the sequence of consonants. Thus, we predicted that learners trying to identify words in a continuous stream would recognize the consonantal tiers better than the vocalic tiers.

Experiment 3 tested this prediction. After exposing participants to a stream that induced word segmentation, we tested memory for items containing intact consonantal tiers but changed vocalic tiers (*CT words*) or intact vocalic tiers but changed consonantal tiers (*VT words*). If participants extract word roots on the basis of the consonants in the speech stream, they should prefer items with their consonantal tier intact.

EXPERIMENT 3

Method

Participants

Fourteen French college students participated.

Stimuli and Procedure

The materials were prepared following the acoustic and randomization constraints of Experiments 1, 2a, and 2b. To construct the continuous speech stream, we used nine words with equally high TPs between adjacent and nonadjacent syllables, as well as between consonants and vowels (1.0 within and .5 between words, respectively; see Table 5).

For the test, 18 new trisyllabic items were constructed (Table 5). None of them had appeared in the familiarization sequences, but 9 had intact consonantal tiers and 9 had intact vocalic tiers. In the test phase, participants were presented 36 pairs, all containing one CT word and one VT word, in an alternative forced-choice paradigm as in the previous experiments. Each test item of one type was presented with four different items of the other type.

TABLE 5
Stimuli Used in Experiment 3

Familiarization			Test	
Consonantal tier	Vocalic tier	Words	CT words	VT words
p_ʀ_g_	_ɑ̃_e_o	pɑ̃ʀego	peʀogɑ̃	dɑ̃kepo
p_ʀ_g_	_u_i_ɛ̃	puʀigɛ̃	piʀɛ̃gy	dukimɛ̃
p_ʀ_g_	_y_ɔ̃_a	pyʀɔ̃ga	pɔ̃ʀagɑ̃	dykɔ̃pa
b_d_k_	_ɑ̃_e_o	bɑ̃deko	bedoku	lɑ̃tebo
b_d_k_	_u_i_ɛ̃	budikɛ̃	bidɛ̃ky	lutipɛ̃
b_d_k_	_y_ɔ̃_a	bydɔ̃ka	bɔ̃dakɑ̃	lytɔ̃ba
m_l_t_	_ɑ̃_e_o	mɑ̃leto	melotu	ʀɑ̃gemo
m_l_t_	_u_i_ɛ̃	mulitɛ̃	milɛ̃ty	ʀugibɛ̃
m_l_t_	_y_ɔ̃_a	mylɔ̃ta	mɔ̃latɑ̃	ʀygɔ̃ma

Note. None of the test items appeared during familiarization. The CT words kept the consonantal tiers of familiarization words intact, and the VT words kept the vocalic tiers of familiarization words intact.

Results and Discussion

Participants preferred CT words over VT words (63.7 ± 10.5% mean preference for CT words), $t(13) = 4.9, p \leq .0001, d = 1.29$ (Fig. 2e). It should be noted that in order to identify words in the streams of Experiment 3, learners had the opportunity to use consonants, vowels, or syllables, because TPs for the three classes were exactly the same. Even so, it was the computation of consonant TPs that won out. Had participants extracted items on the basis of vowel TPs, they would have preferred the items with intact vowel structure in the test phase. Alternatively, had they based their segmentation strategy on syllable TPs, they would have responded at random, because in the test pairs, the syllable structure of both items differed from that of the familiarization words. Instead, participants preferred unheard items with intact consonantal tiers. This result is consistent with the hypothesis that in on-line word-segmentation tasks, learners are capable of extracting not only words, but also word roots identified by their consonantal structures, as in Semitic languages.

BASELINE FREQUENCY AND EXPERIMENTAL FREQUENCY

To assess the influence of previous exposure to French on our data, we queried *Lexique* (New, Pallier, Ferrand, & Matos, 2001), a database containing frequency values computed on 31 million written French words, to estimate the frequency of the consonantal and vocalic tiers of our test items using the frequencies of the phonologically different words containing them. Tables 6 and 7 present the tiers' mean frequencies per million occurrences. The values range from extremely low to zero. As a written corpus heavily overrepresents low-frequency items compared with their actual occurrence in speech, such data inflate the possible effects of prior exposure to French. Even so, in Experiments 1, 2a, 2b, and 2c, any tendency for a possible effect of prior experience with French on our experiments would run counter to our hypothesis. For the test items in Experiment 3, the consonantal tiers had higher frequency values in *Lexique* than the vocalic tiers did, but the advantage was at most 8 occurrences per million. Could such a difference have led participants to selectively concentrate on consonants and disregard vowels? We find that very difficult to believe. The number of instances of the vocalic tiers in the familiarization stream of Experiment 3 was larger than the number of instances of the consonantal tiers that participants are likely to encounter during their lifelong experience with French.

TABLE 6

Mean Frequency of Occurrence per Million and Number of Phonologically Distinct Words Sharing the Tiers of the Test Items in Experiments 1 and 2

	Test items			
	Words		Part-words	
Experiment	Frequency	Number	Frequency	Number
1 (consonantal tier)	2.5	7	4.5	36
2a (vocalic tier)	0.13	6	0.78	9
2b (vocalic tier)	1.4	8	0.0	0
2c (vocalic tier)	0.0	0	2.2	1

Note. The t tests of mean differences, where applicable, were not significant, $p \geq .05$ for Experiment 1, $p \geq .08$ for Experiment 2a. Also, the mean frequency of the consonantal tiers in Experiment 1 was not significantly different from the mean frequency of the vocalic tiers in Experiment 2a, $p \geq .36$.

TABLE 7

Mean Frequency of Occurrence per Million and Number of Phonologically Distinct Words Sharing the Tiers of the Test Items in Experiment 3

	Test items			
	CT words		VT words	
Tiers	Frequency	Number	Frequency	Number
Consonantal	2.5	7	8.0	12
Vocalic	0.0	0	0.7	1

Note. CT words kept the consonantal tiers of familiarization words intact, and VT words kept the vocalic tiers of familiarization words intact. A t test of the difference in mean frequency of the consonantal tiers in CT words and VT words was not significant, $p \geq .36$.

GENERAL DISCUSSION

For a general-purpose statistical device, vowels and consonants are equivalent if their statistical relations are equivalent: Their statistics entirely determine their informational content. However, our French participants could use TPs between consonants, but not between vowels, to discover potential words in continuous speech.

As many languages possess a wide repertoire of consonants and relatively few vowels, it could be argued that it is natural for consonants to have a more central role in speech analysis. However, French enjoys a remarkable balance between vowels and consonants (17 consonants and 16 vowels). No prior numerical asymmetry could lead French listeners to adopt a strategy favoring computations of consonants over computations of vowels, yet the participants in our experiments clearly did so. This asymmetry is especially relevant considering that vowels are acoustically more salient than consonants (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996); hence, if anything, the odds should be in favor of vowels.

Vowels can became salient enough to serve as a basis for segmentation, but only when enhanced by nonprobabilistic concurrent processes, such as when a stream contains long stretches of immediate repetitions of the same vocalic pattern (Newport & Aslin, 2004, and our Experiment 2c). When such additional factors are controlled, vowels do not seem to be open to the same computations that are easily performed on consonants.

Our results suggest not only that consonants are exploited to find words in continuous speech, but also that consonants are independently represented. In Experiment 3, participants preferred words that had intact consonant structures but changed vowel structures over words that had intact vowel structures but changed consonant structures. This suggests that participants extracted words by their consonantal tiers, rather than by their syllable sequences.

What could explain the predominance of consonants in an on-line task of word segmentation? Nespor et al. (2003) have argued that consonants preferentially contribute to lexical processing of words, whereas grammatical variations rest mostly on vocalic properties. On the basis of these observations, we proposed that the task of interpreting the lexicon falls more on consonants than on vowels.

Our present results are consistent with this hypothesis. We suggest that the vowel-consonant asymmetry we have documented depends on the different roles of vowels and consonants in language. If consonants serve mainly to individuate words, whereas vowels tend to carry grammatical information, then extraction of relevant information from speech for purposes of speech segmentation will improve if general-purpose computational resources, such as TP computations, focus on consonants. Blocking computations over vowels, which are not as crucial for word identification, will greatly reduce noise in the system.

However, this argument applies only within a linguistic system; consonants and vowels are linguistic objects to begin with. In organisms without language, such a constraint on computation would not make sense. As it turns out, nonhuman primates listening to a continuous stream of speech do precisely the opposite of human language learners. Exposed to such a stream, tamarins can compute TPs between vowels, but not between consonants (Newport, Hauser, Spaepen, & Aslin, 2004). It thus appears that when vowels lose their linguistic role and become acoustic objects, general learning mechanisms can capture the regularities among them just as well as among any other objects, and when consonants lose their role in word individuation and become hardly distinguishable noises, the animal perceptual system filters them out. Only when a language module exists do consonants and vowels reverse their natural order of saliency.

A crucial issue in cognitive science is to identify the scope and limits of human statistical abilities. We have found that the specific linguistic role of representations severely constrains whether such representations are open to statistical inspection. Language seems to orchestrate where and when general learning mechanisms can deploy their power.

Acknowledgments—The authors contributed equally to this research, which was funded by the grant "European Science Foundation Eurocores Origin of Man, Language and Languages" to Marina Nespor and Jacques Mehler; by Human Frontier Science Project Grant RGP 68/2002; and by Friuli-Venezia Giulia. We thank Michele Nathan for her very careful reading of the manuscript.

REFERENCES

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*, 428–430.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*, 292–295.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In H.T. Bunell & W. Isardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing*. Retrieved March 16, 2004, from http://www.asel.udel.edu/icslp/cdrom/icslp96.htm

Fiser, J., & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458–467.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis*, *2*, 23–68.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

McCarthy, J. (1985). *Formal problems in semitic phonology and morphology*. New York: Garland.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In L.M.K.D. James (Ed.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116). Mahwah, NJ: Erlbaum.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 221–247.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, *101*, 447–462.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162.

Newport, E.L., Hauser, M.D., Spaepen, G., & Aslin, R.N. (2004). Learning at a distance: II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*, 85–117.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments, & Computers*, *29*, 322–327.

Peña, M., Bonatti, L.L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421–435.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Saffran, J.R., Newport, E.L., Aslin, R.N., & Tunick, R.A. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shukla, M., Nespor, M., & Mehler, J. (2004). *Grammar on a language map*. Manuscript submitted for publication.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.

(Received 4/30/04; Revision accepted 7/1/04)