SPECIAL ISSUE: ORIGINAL ARTICLE

THE "SOUL" OF LANGUAGE DOES NOT USE STATISTICS: REFLECTIONS ON VOWELS AND CONSONANTS

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

([1]Cognitive Neuroscience Sector, International School for Advanced Studies-SISSA-ISAS, Trieste, Italy; [2]Universidad Catolica De Chile, Santiago De Chile, Chile; [3]University of Ferrara, Ferrara, Italy; [4]University of Paris, Paris, France)

Abstract

This paper reviews studies of language processing with the aim of establishing whether any type of statistical information embedded in linguistic signals can be exploited by the language learner. The constraints as to the information that can be so used, we will argue, should be used to inform theories of language acquisition.

We present two experiments with their respective controls. Both show that consonants (Cs) are much more suitable than vowels (Vs) to parse speech streams using statistical dependencies. These experiments use streams composed of items in which statistical information is carried either by the sequence of consonants or by the sequence of vowels. Both kinds of items are simultaneously present is the speech stream but, crucially, their overlap is only partial. Since the location of dips in transitional probabilities (TPs) between adjacent syllables differ for the first and the second type of sequences, we can explore whether consonants and vowels are equally efficient segments to parse signals. Our results show that "consonant words" (CW) are significantly preferred over "vowel words" (VW).

We discuss the implication of our results for models of language acquisition.

Key words: language acquisition, vowels, consonants

Introduction

John Marshall's contribution to Cognitive Neuropscyhology spans over many years and many areas. To each area he investigated, John Marshall brought uncommon lucidity and scholarship. He left it to others, however, to make progress in the study of developmental neuropsychology, an area that is still trying to take off. Lenneberg (1967) had argued that exploring a normal population of growing infants, instead of patients, may contribute to improve our understanding of the relation of the emerging functions to brain growth. Initial expectations were not entirely successful. As we shall see below, behavioral explorations have taught us that the emergence of functions does not arise from a blank slate. Yet we are still hoping to establish a deeper understanding of how the emergence of function is related and controlled by the changes in the growing brain. But it was only after behavioral techniques had established a list of basic cognitive precursors in the very young infant that it became possible to think of applying imaging methodologies and other procedures deriving from neuroscience, genetics and more recently animal models to ask questions about the biological foundations of mental capacities. Thus, it is after the hard work carried out by pediatricians, behavioral scientists, cognitive psychologists and linguists that the need for a connection to brain changes began to be strongly felt.

In this paper we first describe summarily some behavioral studies of language acquisition. Second, we discuss briefly the contrast between theorists who believe that language acquisition is just piloted by the distributional richness of the information that speech signals afford and theorists who think that to explain grammar acquisition one requires much more than just studying the distributional properties of speech. Third, we present new experimental studies that ground the contrasting and highly important role played by vowels and consonants during language processing. Finally we show how the observed results might tilt the balance between the two aforementioned theoretical beliefs.

Psychologists and linguists working at the beginning of the twentieth century adopted all encompassing learning theories, which supposedly were capable to explain the acquisition of all skills in all vertebrates (e.g., Skinner, 1957). Those theories sealed the adoption of biologically implausible proposals that became a standard for over one-half a century. Under Chomsky's (1965, 1980) influence, a far more realistic biological perspective began to change this picture. In particular, Chomsky (1965, 1980) argued that humans are endowed with universal linguistic principles and a series of parameters – conceived as a set of binary switches – that must be set to suit the grammar of the particular language the infant is acquiring. Unsurprisingly, Chomsky's (1965, 1980) theory had a great influence in triggering the boom of experimental studies meant to discover infants' predispositions and mental endowment. Those studies show that babies organize their sensory inputs on the basis of primitives, of which some characterize all vertebrates and others are available only to human brains. Spelke et al. (1992) in a foundational paper, argues that infants at birth are not submerged in the famous Jamesian *big blooming buzzing confusion*. Rather, under certain conditions, they actually may see objects when they open their eyes. Vision is possibly one of the most basic input modules leading some to surmise that learning to see objects is hardly necessary; objects in the world may become immediately available to the infant whose representation, as a consequence, becomes fairly "realistic".

In a parallel way, in the auditory domain neonates are endowed with a large number of abilities and preferences. One of the first experimental studies with very young infants focused on their ability to discriminate speech segments either present or absent from the language in their environment, see Eimas et al. (1971) and for a large number of other studies (Jusczyk, 1997; Mehler and Dupoux, 1994). For studies of perception in somewhat older infants, see Werker and Tees (1984) and Kuhl et al. (1992) among others.

Investigators of speech perception were using a framework similar to that of Spelke et al.'s (1992) when they conjecture that "... infants represent objects and reason about object motion in accord with two constraints on the behavior of material bodies: 'continuity' (they don't jump from one place to another without traveling through the connecting path) and 'solidity' (no two distinct objects coincide in space and time)". However, although Spelke et al.'s (1992) notions are intuitively sound and mandatory for vision, they have to be adapted to acoustic objects. Indeed, the representation of acoustic objects is, in all likelihood, determined by properties that characterize the hearing apparatus rather than the eyes. As a matter of fact, language, as conveyed by speech, is a basic auditory object that has become crucial to the human species. How do infants determine which of the multiple sounds in the environment is a speech stimulus to begin with? If the infant has not yet identified the stimuli that carry speech signals, it is unlikely that s/he will learn the phonological properties of the language of exposure.

One of the first conjectures Mehler et al. (1988) made was that when the neonates' brain is activated by speech, attention is heightened and directed towards it. Thus, the conjecture is that speech stimuli become the database upon which the human brain, even at an early age, acquires some characteristic properties of the maternal language. Johnson and Morton (1991) offered a related but more detailed hypothesis to explain how the infant builds a repertoire of known faces, see Johnson and Morton (1991).

Mehler et al. (1988) showed that neonates familiarized with utterances drawn from a given language and tested with utterances from another language are capable of discriminating the switch, in specific cases. Infants may even react to a switch when neither of the languages is the maternal one. Obviously, infants are not born knowing the distinctive properties of all natural languages nor can they compute these after only a few minutes of familiarization. What property of the utterances with which it has been familiarized does the infant extract that enables him/her to detect a language switch? Both Nazzi et al. (1998) and Ramus et al. (1999) showed that infants fail to react to a switch between two languages that belong to the same rhythmic class although they have no difficulty in discriminating two languages that are of a different rhythmic class. To draw such a conclusion Ramus et al. (1999) gave a formal definition of linguistic rhythm based on the proportion of time – per representative utterance – occupied by vowels and on the variability of the intervocalic intervals.

This characterization of linguistic rhythm made it possible to demonstrate that the language pairs the infant discriminates are languages that belong to two different rhythmic classes of the type first observed in classical phonological studies (Pike, 1945; Abercrombie, 1967; Ladefoged, 1975). In contrast, infants fail to discriminate switches between languages that belong to the same rhythmic class. Since the Ramus et al. (1999) representation of the rhythmic space contained only nine languages, most of which were related, in subsequent work, many new languages from a diversified geographical and linguistic origin have been added see Shukla et al. (in preparation). In addition, we recently conjectured that rhythm correlates with certain syntactic properties, a notion that was already present in Nespor and Vogel (1986).

In summary, infants are capable of categorizing vowels and consonants, a categorization which allows them to compute rhythmic patterns; such patterns may provide information about the more abstract properties of language.

Empirical data obtained from brain studies with adults suggest that vowels and consonants are also differentially processed. In fact, syllabic discrimination task relying on detection of a change in consonants, but not one with a change in vowel, becomes very difficult after a cortical stimulation over a small area of the superior temporal gyrus in the LH (Boatman, 1997). Likewise, some brain lesions have been associated with errors in either the production of vowels or that of consonants. Given that the errors do not correlate with the levels of sonority of the targets, different phonological brain network for vowels and consonants have been proposed (Caramazza et al., 2000).

Statistics and Perceptual Constraints

Spinoza in his *Compendium Grammaticus Lingua Hebraeae* (1677) states that “La lettre est le signe d’un mouvement de la bouche, mouvement différent suivante l’origine du son émis”[1]. It seems obvious that the word “lettre” translates into the word “consonant” since in the same line Spinoza (1677) goes on to write that “les Hébreux disent que ‘les voyelles sont l’âme des lettres’… et les lettres sans voyelles sont des ‘corps sans âme’”[2].

The whole paragraph reads as follows:

“Ce sont les accents, leur mélodie entraine en mouvement derriere eux, comme une armée derriere le Roi, les lettres et les voyelles. Les lettres, ce sont le corps; les voyelles, sont l’âme; toutes suivent la marche des accents et s’arretent en meme temps qu’eux”[3]. We quote from the French version entitled *Abrégé de Grammaire Hébraïque*, published in Paris in 1968. Spinoza’s (1677) insights are being vindicated by numerous recent empirical reports, as we shall see below.

The graphic system of Hebrew may have inspired Spinoza (1677) to ponder about the basic functional/categorical division of consonants and vowels. It is uncanny that an author writing in the second half of the sixteenth century would have foreseen some recent discoveries in Cognitive Psychology and Neuropsychology. His second passage quoted above reminds us of the distinct categorical nature of vowels and consonants, a realization that is also currently a matter of active research.

Spinoza’s (1677) proposal can be taken somewhat further to explore whether the language user can exploit as efficiently the information-theoretical properties carried by vowels and those carried by consonants. As an aside, we believe that Spinoza (1677) would have scoffed at those hard headed who would have dared to claim that humans compute statistics on “soul” matter. In contrast, we suspect that he might have been more tolerant to claims that consonants are better than vowels to compute information theoretical dependencies – like transition probabilities.

How Cs and Vs Differ Linguistically

There are many reasons to consider that consonants (Cs) and vowels (Vs) are separate linguistic categories with different functions in the linguistic system. Nespor et al. (2003) conjectured that consonants bear the main burden of distinguishing lexical items, whereas vowels principally provide cues to grammatical information. In most linguistic systems, Cs outnumber Vs and the prototypical repertoire consists of roughly 20 Cs and 5 Vs. Systems with a similar number of Cs and Vs are rarely attested, e.g., Swedish that has many phonemes in both categories, and Rotokas (a language of Papua New Guinea) that has very few (Maddieson, 1984). The common higher proportion of Cs compared to Vs is one of the many properties that point to Cs as being best suited to distinguish lexical items as compared to Vs.

The distinctive power of vowels is further reduced through a number of phonological phenomena. In many languages, e.g. English, vowels in unstressed position are reduced to *schwa*, the unstressed vowel pronounced with the mouth in rest position, for example, *a* in *above* or *e* in *sicken*. In other languages, e.g. Portuguese or Italian, the vocalic repertoire in unstressed position is reduced with respect to the repertoire in stressed position. In yet other languages, e.g. Turkish, vowels harmonize throughout a word.

Consonants tend to become more, rather than less, distinct. They tend to disharmonize in many languages, e.g. Japanese (Itô and Mester, 1986) or Arabic (McCarthy, 1985); that is, if several word internal Cs share the same feature, one changes its value. The distinctive power of Cs is particularly obvious in Semitic languages, where the roots of words consist exclusively of Cs, namely those that Spinoza (1677) considered as the *letters*. A word thus has a consonantal root and the vowels that separate the consonants, the *soul* in Spinoza’s (1677) terms, generate different words or word forms. In addition, even if both Vs and Cs vary greatly, only Cs tend to be perceived categorically (Lisker and Abramson, 1964). Indeed, variations within a category tend to be neglected while discrimination is much better close to the boundary between the categories (Liberman et al., 1957).

Both Vs and Cs alternate, but while Cs tend to alternate in quality, as we have seen, Vs tend to alternate in quantity. As elements that carry stress and accent information, vowels are the main carriers of prosody, which can be characterized by alternation of more and less prominent elements (Yip, 1988). In most languages two primary word stresses in adjacent syllables are prohibited and if words are put together in such a way that this configuration occurs, one of several different phenomena applies to reestablish alternation (Liberman and Prince 1977; Nespor and Vogel 1989). Through prosody, vowels signal both universal syntactic properties, such as the edges of constituents of certain types, and properties that vary across languages, such as different word orders Nespor and Vogel (1986). We will now turn to the different roles of Vs and Cs in basic psycholinguistic processing.

---

[1]The letter is the sign of a mouth movement, movement that differs according to the sound produced.

[2]The Jewish say that the vowels are the soul of the letters (and the letters are) bodies without soul.

[3]Accents and their melodies surge behind them, as an army, the letters and the vowels behind the King. The letters are the body; vowels are the soul; all of them follow the march of accents and stop at the same time as them.

Evidence that Cs and Vs Are Not Used for The Same Purposes

Transition probabilities (TPs) between segments provide essential information about some properties of the surrounding language, a notion that was first proposed by Hayes and Clark (1970). TPs between adjacent syllables are not only useful to discover phonotactic regularities: infants as young as eight months can use these to segment continuous, monotonous, meaningless streams of speech into its constituent "words" (Saffran et al., 1996). After listening for a few minutes to such a stream, infants reacted differently to "words" that had high TPs between all the syllables, as compared to "part-words" that contained a dip in TP between the last syllable(s) of one item and the initial syllable(s) of the next one. Saffran et al. (1996) elegant experiments demonstrate that while most speech sequences that surround infants are continuous, the infant will have little trouble in breaking up such sequences into constituent words using TPs. For a contrasting viewpoint, see Yang (2004).

Newport and Aslin (2004) went even further and showed that if one considers Cs and Vs as different categories, and if one believes that each category might be represented in a separate sequence or linear tier, then one may think of a CVCVCVCV-like stream as yielding in one representation the aligned Cs as adjacent to one another and in another one the aligned Vs as adjacent to one another, see Goldsmith (1990). Newport and Aslin (2004) established experimentally that adult participants could segment the streams using either the C sequence or the V sequence if the TPs between the elements of either the one or the other representation were sufficiently informative. The streams thus consisted of words belonging to either families that shared the Cs while the Vs varied or to families that shared the Vs while the Cs varied. Indeed, participants in the Newport and Aslin (2004) study were able to segment the speech stream when the syllables had TPs that were identical both within and between "words" and either the Cs or Vs linear sequence had a TP dip at the end of the constituent "words".

In our laboratory we carried out similar studies. In particular, given the theoretical interest in the categorical divide between Cs and Vs, Bonatti et al. (2005) carried out similar studies as the ones by Newport and Aslin (2004). We found that French participants behaved like those in the Newport and Aslin (2004) experiments only in relation to the C sequences. Indeed, when a linear sequence of consonants defined "words" (realized with varying Vs), the participants rated new "words" as being familiar if the linear sequence of Cs remained intact. However, with the corresponding Vs experiment we observed results that differ from the ones reported by Newport and Aslin (2004). Indeed, while the Rochester participants performed alike with Cs as they had with Vs, our participants only succeed with the informative Cs experiment but failed with the equally informative Vs experiment. In fact, in one of Bonatti et al.'s (2005) experiments it was possible to test how well the C-sequences are recognized as compared to the V-sequences. Unmistakably, our participants were more successful with the C-sequences.

There are several differences in the materials used by the two groups that might explain the contrasting results. Bonatti et al.'s (2005) original streams never allowed repetitions of two items belonging to the same C-family or V-family. In contrast, Newport and Aslin (2004) (replicated in Bonatti et al., 2005, Experiment 3) had to use repetitions given that they were using a more modest number of families of words. Indeed, the differing number of families used in the two studies might also be a factor that explains the differences between the two experiments.

To buttress the significance of our results we present new results recently carried out to evaluate simultaneously the functional validity of Vs and Cs TPs to segment the speech streams with which participants are confronted. This entails simultaneously pitting the C-families against the V-families. Below we present two experiments that should clarify the ease with which the two categories of speech sounds can serve as input to carry out statistical computations in view of speech segmentation.

Segmenting Continuous Speech Using Consonants or Vowels

Neither Newport and Aslin (2004) nor Bonatti et al. (2005) used a design to evaluate directly the functional validity of consonant and vowels as categories to segment continuous speech. Below, we present a series of studies that confirm that native French speakers segment continuous speech using the TPs over consonants but not over vowels.

We present two experimental studies designed to assess in a more direct manner than has previously been done (see Bonatti et al., 2005 and Newport and Aslin, 2004), the relative importance, for the purpose of speech segmentation, of consonants and vowels. The two categories are, in principle, equally suitable phonemes to compute distributional information to segment continuous speech streams. Indeed, the purpose of the experiments we present below is to evaluate how efficient sequences of Cs are in comparison to sequences of Vs to parse speech streams. Of course, to achieve our aim it is necessary to design such an experiment making the TPs between syllables uninformative about word boundaries, yet capable of being simultaneously informative through the TPs in V-, and C-sequences. As will be illustrated below, a given C-sequence will generate a number of consonant-words (CW) by changing the vowels with which the syllabic stream is realized. Likewise, vowel-words (VW) are generated by fixed sequences of Vs but changing the Cs with which the syllabic stream is realized. Indeed, our design made it possible to exploit the TPs between consonantal sequences and/or TPs between vowel sequences allowing us to then test participants with pairs of VWs and CWs that were heard equally often in the stream. This makes it possible to estimate the functional validity of vowels as compared to consonants as categories best suited to carry out TP computations for the purpose of segmentation. In order to achieve this aim it was also necessary to have two control streams that provide no consistent statistical information to detect words on the basis of consonant, vowel or syllable TPs. Participants familiarized with the control stream were offered the same pair of test items to estimate their spontaneous preferences.

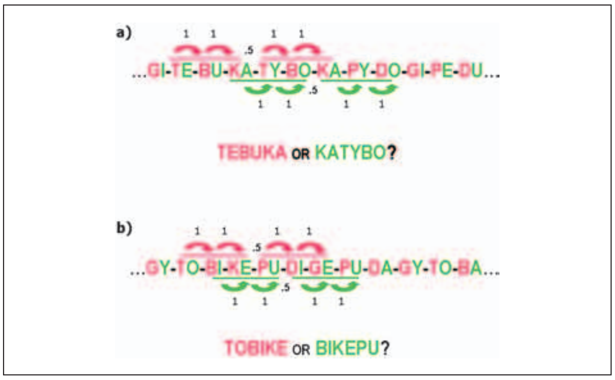

Fig. 1 – *(a) A small stretch of the familiarization stream is depicted with consonants in red and vowels in green for Experiment 1. Syllables are separated by hyphens. The number over the arrows in red indicate the TP between adjacent consonants and the numbers below the green arrows that of the vowels. The numbers between arrows give the TP between CWs or VWs. The words below the familiarization stream present one of the 32 test pairs. (b) A small stretch of the familiarization stream is depicted with consonants in red and vowels in green for Experiment 2. Syllables are separated by hyphens. The numbers over the red arrows indicate the TP between adjacent consonants and the numbers below the green arrows that of the vowels. The numbers between arrows give the TP between CWs or VWs. The words below the familiarization stream are one of the 32 test pairs.*

Experiment 1

Materials

Two continuous streams were assembled using 12 CV syllables arranged to obtain sixteen words. In Experiment 1, a planned concatenation was arranged such that the transition probabilities between syllables did not provide consistent TP cues to segment "words". Instead, the concatenation was arranged so as to include both high TPs between successive Cs within CWs and between successive Vs within VWs, but with dips in TPs situated at the respective word edges, as displayed in Figure 1a. Eight words, hereafter referred to as "consonant words" (CW) are characterized by a TP of 1.0 between all its successive consonants. Another eight words are the "vowel words" (VW) characterized by TPs of 1.0 between all its successive vowels. Four CWs contained the consonantal sequence "t-b-k" and the other four contained the sequence "p-d-g". Likewise, the stream contained eight VWs, four which contained the "a-y-o" vowel sequence and four that contained the "i-e-u" sequence. The CWs and VWs used are shown in Table I.

The two C-sequences, i.e., "p_d_g_" and "t_b_k_" were pseudo-randomly concatenated respecting the following restriction: a given CW cannot immediately follow itself. In fact, a particular given CW could be followed by two other CWs from the same sequence and by two CW from the other sequence. Another restriction to concatenate the syllables was that two CWs could not follow one another if they shared more than two consecutive identical syllables. Thus, a CW like *pydogi* could be followed by *pedugi, peduga, tebuka and tebuki.* Identical constraints hold for adjacent VWs. Due to the above constraints, streams with eight CWs and eight VWs were generated ensuring that the CWs and the VWs overlap only partially. All the adjacent TPs between consonants within CWs were equal to 1. The TP between the last C of a CW and the first C of the following CW was kept at .5. Similarly, the TPs between vowels in the VWs was equal to 1, and the TP between the last V of a VW and the first V of the following VW was equal to .5. Figure 1a illustrates the TP distributions described above. Notice that the boundaries of the CW fall inside a VW and vice versa. In CWs the TPs between adjacent syllables was equal to 1 between first and second syllable and equal to .5 between second and third syllable. In VWs, the TPs between adjacent syllables were equal to .5 between first and second syllable and equal to 1 between the second and third syllable.

Speech streams were prepared using MBROLA a text to speech software implementing the synthesis

TABLE I

Material used to construct the familiarization stream and test items used in Experiment 1 (in sampa diphones)

Consonantal sequence	Vocalic sequence	"Consonant-words"	"Vowel-words"
"p_d_g_"	"a_y_o_"	pydogi	katybo
		peduga	kapydo
		pydoga	gapydo
		pedugi	gatybo
"t_b_k_"	"i_e_u_"	tebuki	gipedu
		tyboka	gitebu
		tebuka	kitebu
		tyboki	kipedu

with French diphones. The duration of syllables in the stream was 232 msec and the duration of the familiarization stream was 14 minutes.

Participants

Fourteen French adult speakers were individually tested in a sound proof booth using high-quality headphones. Volunteers were paid 3 euros for participating in the experiment.

Test

After familiarization, participants were asked to choose one of two items presented in a two alternative forced choice task. Thirty-two test pairs, each containing a CW and a VW with 500 msec ISI, were presented auditorily. The positions occupied by CWs and VWs across trials were counterbalanced. The duration of the CWs and VWs was 696 msec.

Results

Figure 2 shows the results for Experiment 1. We found a significant preference for CWs over VWs [75.4 ± 24.5; t (13) = 3.9; p < .002].

To assess whether the preference for CWs is due to the familiarization rather than to the spontaneous preference for some of the CWs we tested a new group of fourteen participants with streams using 128 different three-syllabic items thus lowering TPs between all units: consonants, vowels and syllables. These three-syllabic items were constructed using the same set of syllables as in Experiment 1, but combined at random. At the end of a 14 min. familiarization with the random stream, participants were exposed to the same test pairs as those in Experiment 1. Figure 2b shows the results of this control. We found a significant preference for VWs over CWs [39.4 ± 12.4; t (13) = – 3.1; p < .009] suggesting, if anything, a spontaneous preference for VWs.

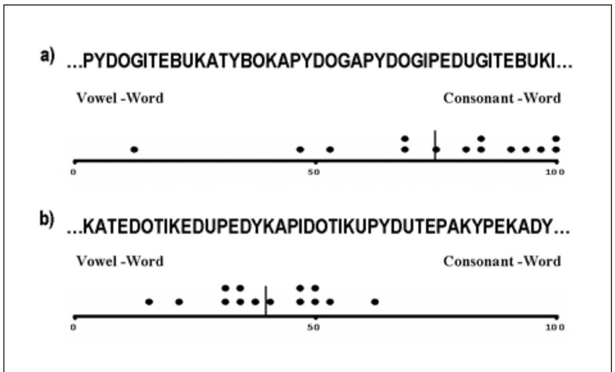

Fig. 2 – *The results in a) refer to Experiment 1 while those in b) refer to the control. To the right of a) and b) the letters represent a stretch of the syllables in the familiarization streams. The results of Experiment 1 show a preference for CWs. Each black dot represents the mean test score for one participant. The horizontal line represents the percentage of choice of word-types. The vertical line represents the mean group performance.*

The results obtained suggest that CWs are recognized significantly better than VWs despite the fact that participants had a spontaneous preference for VWs. Moreover, since the CWs had TPs of 1.0 between the Cs and the VWs between the Vs, this clear preference might be uncovering the fact that consonant sequences might be more effective as a cue to segmentation than vowel sequences. To ground this conjecture, however, it is necessary to add another experiment. Indeed, as we mentioned above, in Experiment 1 the TPs between adjacent syllables of CV and VW was asymmetrically distributed. In fact, in CWs the TP between adjacent syllables is reduced in the transition from the 2nd to the 3rd syllable while in VWs the TP between adjacent syllables is reduced between the 1st and the 2nd syllable. To explore whether such an asymmetry could explain our results, we run Experiment 2 by reversing the location at which the reduction of TPs between syllables occurs.

Experiment 2

Familiarization

The continuous speech streams used in Experiment 2 was identical to the one used for Experiment 1, with one exception: here the concatenation of the twelve syllables generated VWs with a TP of 1.0 between syllable 1 and syllable 2 while the TP between syllable 2 and syllable 3 was .5. Of course, the values of the TPs between the syllables are reversed for the CWs. The materials are presented in Table II.

The two VW families, "a_y_o_" and "i_e_u_" were pseudo-randomly concatenated. In fact, a given VW was concatenated with only two VWs of the same family and with two VWs of the other family. For instance *dagyto* can be followed by *bikepu, biketu*, *bakyto* or *bakypo*. Due to these constraints, 8 three-syllabic CWs are generated partially overlapping with the VWs. The TPs between vowels was equal to 1 within VWs and .5 at the VWs boundary. Similarly, the TP between consonants was equal to 1 within the CWs and .5 at the boundaries of the CWs. In contrast to the distribution of syllabic TPs we had in Experiment 1, in the present

TABLE II
Materials used in Experiment 2

Vocalic sequence	Consonantal sequence	"Vowel-words"	"Consonant-words"
"a_y_o_"	"p_d_g_"	dagyto bakypo dagypo bakyto	tobaky tobike tubike tubaky
"i_e_u_"	"t_b_k_"	biketu digepu bikepu digetu	pudige pudagy podagy podige

Experiment 2 the TPs between adjacent syllables for VWs was equal to 1 between first and second syllable and it was equal to .5 between second and third syllable (Figure 1b). In contrast, the TPs of the adjacent syllables for CWs were equal to .5 between the first syllable and the second syllable and 1 between the second and the third.

Test

Test phase was identical to that in Experiment 1.

Participants

Fourteen adult speakers of French completed Experiment 2.

Results

Figure 3a shows the results obtained for Experiment 2. We found a significant preference for CWs over VWs [84.4 ± 20.9; t (13) = 6.1; p < .00003].

In order to verify that these results could not have arisen due to spontaneous preferences for the test items we used, we ran a very similar control as that of Experiment 1. Indeed, we used the same stream as in the previous control but the test items of Experiment 2. Fourteen adult French speakers were tested in this control experiment. None of the participants had been tested either in Experiment 1 or in its control.

Figure 3 shows the results. There was not a significant preference for either VWs or for CWs [47.5 ± 11.4; t (13) = – .8; p < .4].

Discussion

Taken together the results of Experiments 1 and 2, strongly suggest that when CWs and VWs overlap but do not coincide in a continuous stream of monotone speech, participants are highly consistent in choosing the CWs over the VWs although their frequency is identical. Moreover, these results cannot be accounted for by a spontaneous preference for one or the other type of words. The controls that accompany both experiments ruled out that possibility. Last but not least, in these experiments, syllable TPs were of no use to segment the speech stream.

The results presented above can be compared to those that were presented in Bonatti et al. (2005), which also tried to evaluate the functional role of consonants and vowels in a single experiment. After having grounded the fact that high TPs between Cs yield successful segmentation of continuous speech streams while high TPs between Vs do not, Bonatti et al. (2005) went on to test in their Experiment 4, memory for C-sequences constituting CWs as compared to V-sequences constituting VWs. The authors used nine words, with equally high TPs between adjacent and non adjacent syllables, consonants and vowels, that is TP = 1 within words and TP = .5 between words, respectively. The results clearly showed that when they had to choose between new items with the same C-sequence of CWs and new items with the same V-sequence of VWs, the first was preferred. In order to segment the stream into words in these experiments learners could rely upon consonants, vowels and/or syllables since the TPs for the three classes were identical. In the present experiment the design eliminates the influence of TPs between syllables, thus allowing us to directly compare the preference for either CWs or VWs.

The above results do not mesh well with the results reported in Newport and Aslin (2004) who report that English speaking learners exposed for 21 minutes to streams characterized by either C-TPs or V-TPs could segment on the basis of both.

In a follow up experiment Newport et al. (2004) also compared white-cotton-top tamarin monkey confronted with comparable streams. Interestingly, the monkeys, despite generally testing like human adults in many previous experiments with language stimuli, behaved differently than the students tested previously by Newport and Aslin (2004). Indeed, Tamarins compute the TPs between vowels but not those between consonants (Newport et al., 2004). In Bonatti et al. (2005) we interpreted this surprising result as an indication that the monkeys use vowels as arbitrary acoustic objects and "general learning mechanisms can capture 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".

Another difference between the Newport and Aslin (2004) paper and Bonatti et al.'s (2005) paper consists in the way in which adjacent repetition of

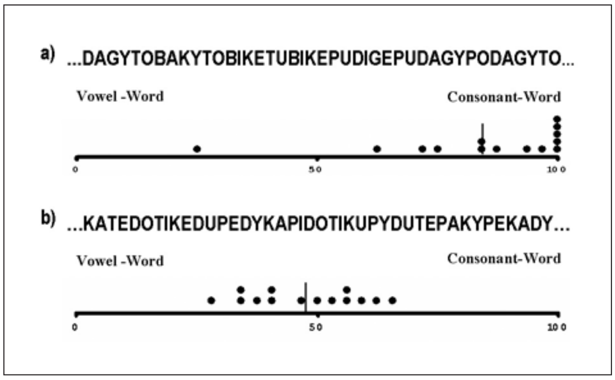

Fig. 3 – *The results in a) refer to Experiment 2 while those in b) refer to the control. To the right of a) and b) the letters represent a stretch of the syllables in the familiarization streams. The results of Experiment 2 show a preference for CWs. Each black dot represents the mean test score for one participant. The horizontal line represents the percentage of choice of word-types. The vertical line represents the mean group performance.*

segmental frame-words are handled. To test the CWs and the VWs, in Experiment 2, Newport and Aslin (2004) used two streams in which they had exactly either two C-families or two V-families to test recognition of CWs and VWs, respectively. This arrangement might have lead to a differential sensitivity to CWs and VWs in the experiments of the two groups. Indeed, Endress et al. (2005) have shown that adjacent repetitions have a very special status since they pop-out in ways that are not seen for most other structural regularities.

The results of our experiments suggest that even when continuous speech contains repetitions of C-sequences as frequent as that of V-sequences, TP computations rely significantly more on C-sequences to discover potential words.

Conclusions

The results presented in this paper demonstrate in a more direct and controlled fashion than previously reported that high TPs between a sequence of consonants in a CW – followed by a TP dip before the sequence of consonants of next provide an efficient cue for language learners to segment an otherwise continuous speech stream. Streams that incorporate high TPs between the sequences of vowels of a VW – followed by a TP dip before the next sequence of vowels constituting a VW do not help language learners to segment speech streams, except in very special cases, like when V-sequences constituting VWs are repeated immediately one after the other. Although high TPs between vowels might have some use under specific circumstances, we have demonstrated in a direct comparison that consonantal sequences are far more effective to segment and recognize items contained in the stream than vowel sequences that appeared equally frequently.

Though our results have been obtained by testing adult volunteers confronted with an unknown artificial language, it is conceivable that consonantal representations mediate also language acquisition in infants. We are currently exploring this possibility and it appears indeed likely that infants, as adults, exploit consonantal sequences crucially more than vocalic sequences.

Our experiments mesh well with Spinoza's (1677) intuitions, which we presented in the first part of this paper. Indeed, Spinoza (1677) had the view that vowels and consonants are radically different categories of speech. Since his classical writings, phonologists have generally accepted that vowels and consonants are different categories and even acoustic phoneticians have attributed to Vs and Cs different roles, see Blumstein and Stevens (1981). While vowels allow speech to be transmitted over rather long distances, Cs modulate Vs to obtain a far larger number of distinctive syllables. Likewise, Caramazza et al. (2000) have observed that one patient with a lesion in the left parietal and temporal lobes plus a small lesion in the right parietal lobe show speech production errors that reflect Vs processing problems while another patient with a lesion localized in left supramarginal, angular and superior temporal gyri has production problems realized in errors in the processing of Cs. A number of control studies have shown that the relative sonority of the Cs is not a parameter that plays a role in the amount of difficulty the first patient presents with the processing of Cs, leading the authors to conclude that different neural mechanisms should be responsible for their processing.

It is interesting to mention that there are a number of languages, like the Semitic languages, that have word families that share some semantic content that appears to be carried exclusively by the consonantal sequence. In contrast, we are not aware of the existence of languages that contain word families that are characterized exclusively by a vowel sequence.

Thus it appears that the contrast between Vs and Cs is not only important for speech processing and most likely for language acquisition, but it also allows us to draw some general conclusions about the role of the two segmental classes in the organization of language.

The fact that linguistic representations constrain the domain over which statistics can be calculated shows that even domain general learning mechanisms may be constrained in a language specific way: distributional properties of speech thus are not sufficient to fully understand how language is processed and possibly how it is acquired.

John Marshall's conception of Cognitive Neuropsychology has influenced our ways of conceiving models of language acquisition. Rather than producing all encompassing theories that can eventually accommodate in more or less *ad hoc* fashion all types of observations, we, like Marshall, try to isolate single components from the complexity of speech to establish how some of these affect in a signal driven fashion different stages of learning.

Acknowledgements. Various parts of the study were supported by the HFSP grant RGP 68/2002, the Regione Friuli-Venezia-Giulia (L.R. 3/98), the McDonnell Foundation, the CEE contract 12778 (NEST) specific targeted project CALACEI and the Progetto FIRB. The research is also supported in the framework of the European Science Foundation EUROCORES program "The Origin of Man, Language and Languages" as well as C.O.F.I.N. 2003-2005.

References

Abercrombie D. *Elements of General Phonetics.* Chicago: Aldine, 1967.

Blumstein SE and Stevens KN. Phonetic features and acoustic invariance in speech. *Cognition, 10:* 25-32, 1981.

Boatman D, Hall C, Goldstein MH, Lesser R and Gordon B. Neuroperceptual differences in consonant and vowel

discrimination: As revealed by direct cortical electrical stimulation. *Cortex, 33:* 83-98, 1997.
BONATTI L, PEÑA M, NESPOR M and MEHLER J. Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16:* 451-459, 2005.
CARAMAZZA A, CHIALANT D, PAPASSO R and MICELI G. Separable processing of consonants and vowels. *Nature, 403:* 428-430, 2000.
CHOMSKY N. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
CHOMSKY N. *Rules and Representations*. New York: Columbia University Press, 1980.
EIMAS PD, SIQUELAND ER, JUSCZYK P and VIGORITO J. Speech perception in infants. *Science, 171:* 303-306, 1971.
ENDRESS AD, SCHOLL BJ and MEHLER J. The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 3134:* 406-419, 2005.
GOLDSMITH J. *Autosegmental and Metrical Phonology*. Cambridge: Blackwell, 1990.
HAYES JR and CLARK HH. Experiments on the segmentation of an artificial speech analogue. In Hayes JR (Ed), *Cognition and the Development of Language*. New York: Wiley, 1970.
ITÔ J and MESTER A. The phonology of voicing in Japanese. *Linguistic Inquiry, 17:* 49-73, 1986.
JOHNSON M and MORTON J. Conspec and conlearn: A two-process theory of infant face recognition. *Psychological Review, 98:* 164-181, 1991.
JUSCZYK PW. *The Discovery of Spoken Language*. Cambridge, MA: MIT Press, 1997.
KUHL PK, WILLIAMS KA, LACERDA F, STEVENS KN and LINDBLOM B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255:* 606-608, 1992.
LADEFOGED P. *A Course in Phonetics.* New York: Harcourt Brace Jovanovich, 1975.
LENNEBERG EH. *Biological Foundations of Language*. New York: Wiley, 1967.
LIBERMAN AM, HARRIS KS, HOFFMAN HS and GRIFFITH BC. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54:* 358-368, 1957.
LIBERMAN M and PRINCE A. On stress and linguistic rhythm. *Linguistic Inquiry, 8:* 249-336, 1977.
LISKER L and ABRAMSON A. A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20:* 384-422, 1964.
MADDIESON I. *Patterns of Sounds*. Cambridge: Cambridge University Press, 1984.
MCCARTHY J. *Formal Problems in Semitic Phonology and Morphology*. Doctoral dissertation. New York: Garland Press, 1985.
MEHLER J and DUPOUX E. *What Infants Know.* Cambridge, MA: Blackwell, 1994.
MEHLER J, JUSCZYK P, LAMBERTZ G, HALSTED N, BERTONCINI J and AMIEL-TISON C. A precursor of language acquisition in young infants. *Cognition, 29:* 143-178, 1988.
NAZZI T, BERTONCINI J and MEHLER J. Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24:* 756-766, 1998.
NESPOR M, PEÑA M and MEHLER J. On the different role of vowels and consonants in language processing and language acquisition. *Lingue e Linguaggio, 2:* 221-247, 2003.
NESPOR M and VOGEL I. *Prosodic Phonology*. Dordrecht: Foris, 1986.
NESPOR M and VOGEL I. On clashes and lapses. *Phonology, 6:* 69-116, 1989.
NEWPORT EL and ASLIN RN. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48:* 127-162, 2004.
NEWPORT EL, HAUSER M, SPAEPEN G and ASLIN RN. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49:* 85-117, 2004.
PIKE KL. *The Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.
RAMUS F, NESPOR M and MEHLER J. Correlates of linguistic rhythm in the speech signal. *Cognition, 73:* 265-292, 1999.
SAFFRAN JR, ASLIN RN and NEWPORT EL. Statistical learning by 8-month-old infants. *Science, 274:* 1926-1928, 1996.
SHUKLA M, NESPOR M and MEHLER J. Grammar on a language map. In preparation.
SKINNER BF. *Verbal Learning and Verbal Behavior*. New York: Appleton-Century-Croft, 1957.
SPELKE ES, BREMLINGER K, MACOMBER J and JACOBSON K. Origins of knowledge. *Psychological Review, 4:* 605-632, 1992.
SPINOZA B. *Compendium Grammatices Linguae Hebraeae (1677).* French Translation *Abrégé de Grammaire Hébraïque*. Paris: Librairie Philosophique J. Vrin, 1968.
WERKER JF and TEES RC. Cross-language speech perception evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7:* 49-63, 1984.
YANG CD. Universal Grammar, statistics or both? *Trends in Cognitive Science, 8:* 451-456, 2004.
YIP M. The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry, 19:* 65-100, 1988.

Jacques Mehler, Settore di Neuroscienze Cognitive, Scuola Internazionale Superiore di Studi Avanzati (SISSA-ISAS), via Beirut 2-4, 34014 Trieste, Italia.
e-mail: mehler@sissa.it

(*Received 18 March 2005; reviewed 29 August 2005; revised 15 September 2005; accepted 10 February 2006*)