ELSEVIER

Cognition 70 (1999) 27–52

COGNITION

Statistical learning of tone sequences by human infants and adults

Jenny R. Saffran[a,*], Elizabeth K. Johnson[b], Richard N. Aslin[a], Elissa L. Newport[a]

[a]*University of Rochester, Rochester, New York, NY, USA*
[b]*Department of Psychology, Johns Hopkins University, Baltimore, MD, USA*

Received 30 July 1998; accepted 23 November 1998

Abstract

Previous research suggests that language learners can detect and use the statistical properties of syllable sequences to discover words in continuous speech (e.g. Aslin, R.N., Saffran, J.R., Newport, E.L., 1998. Computation of conditional probability statistics by 8-month-old infants. Psychological Science 9, 321–324; Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274, 1926–1928; 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., Barrueco, S., 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science 8, 101–195). In the present research, we asked whether this statistical learning ability is uniquely tied to linguistic materials. Subjects were exposed to continuous non-linguistic auditory sequences whose elements were organized into 'tone words'. As in our previous studies, statistical information was the only word boundary cue available to learners. Both adults and 8-month-old infants succeeded at segmenting the tone stream, with performance indistinguishable from that obtained with syllable streams. These results suggest that a learning mechanism previously shown to be involved in word segmentation can also be used to segment sequences of non-linguistic stimuli.

Keywords: Continuous speech; Statistical learning ability; Word segmentation

* Corresponding author. Present address: Department of Psychology, University of Wisconsin at Madison, Madison, WI 53706-1696, USA. E-mail: jsaffran@facstaff.wisc.edu

0010-0277/99/$ - see front matter
PII: S0010-0277(98)00075-4

1. Introduction

Everyday experiences appear to fall naturally into different domains. For example, to a naive listener, an Indian lullaby and an Argentinean tango are more similar than Hamlet's soliloquy and a Bach fugue. Behaviors across domains, such as language and music, are perceived in qualitatively distinct categories, despite common structural attributes (e.g. Lerdahl and Jackendoff, 1983). Furthermore, the neural mechanisms relevant to different domains are often topographically distinct. For example, most aspects of language are processed by the left hemisphere, while processing of pitch structure in music mostly involves the right hemisphere (e.g. Kimura, 1964; Bever and Chiarello, 1974; Peretz, 1987; Zatorre et al., 1992; Anderson, 1994 (as cited in Balaban et al., 1998)), and this pattern of lateralization also appears to be present in young infants (Best et al., 1982; Bertoncini et al., 1989; Balaban et al., 1998).

How might this differentiated or domain-specific end-state arise? The most straightforward answer is that distinct kinds of knowledge are acquired by distinct knowledge-acquisition mechanisms. This view is present, explicitly or otherwise, in theories which range far and wide along other dimensions. For example, modular theories hypothesize that learning is accomplished by mechanisms particular to the domain in question (e.g. Fodor et al., 1974; Chomsky, 1975a; Fodor, 1983). Interestingly, connectionist models suggest a similar conclusion. While the particular architectures and learning algorithms used in any given network are often not designed to solve a particular knowledge-acquisition problem, learning in one domain requires the network to be devoted to that domain henceforth: 'initially, a network could be trained to process physics or linguistic input data. But after learning (say) linguistic data, the same network becomes incapable of learning physics data without undoing all the learning it had achieved for the initial input set. At one level of description, then, networks are just as domain specific as many instances of human learning' (Karmiloff-Smith, 1992, p. 181; see also the discussion of temporal crosstalk by Jacobs et al., 1991).[1]

In the present paper, we ask whether learning in different domains can be at least partly subserved by the same knowledge-acquisition processes (for related discussions of this problem, see also Karmiloff-Smith, 1992; Kelly and Martin, 1994; Elman et al., 1996; Gelman and Williams, 1998). The literature suggests that in some cases, the presence of domain-specific learning abilities is incontrovertible; the acquisition of birdsong is a prime example (e.g. Marler, 1991). In other cases, domain-specific learning abilities are highly implausible. For example, humans commonly acquire categories of knowledge for which they cannot have possibly evolved specialized learning mechanisms, including much of what we learn in school and games such as chess and Go.

The studies reported here consider an intermediate case where the available evidence is unclear. When either a domain-specific or a somewhat more general

[1]This problem might be circumvented by a network charged with the task of acquiring both systems simultaneously.

learning mechanism is plausible, one cannot draw strong conclusions by studying only the domain in question. Instead, one must examine the putative workings of this mechanism across multiple domains. In the present research, we asked whether a learning mechanism which contributes to the acquisition of one aspect of language, word segmentation, can also subserve learning in another domain, the grouping and segmentation of tonal sequences.

2. Word segmentation

One of the initial problems confronting language learners is the continuous nature of fluent speech. Unlike the white spaces available in written text, the speech stream does not contain consistent physical cues marking word boundaries (e.g. Cole and Jakimik, 1980). This raises a difficult problem for infants, who must somehow determine which sequences of sounds are words and which are not. Despite the complexity of this learning task, infants as young as 7.5 months of age demonstrate the ability to extract words from continuous speech (Jusczyk and Aslin, 1995).

Infants are likely to use a number of different types of information in tandem to discover word boundaries, including prosodic cues and silences at the ends of utterances (e.g. Mehler et al., 1990; Jusczyk et al., 1993; Morgan and Saffran, 1995; Aslin et al., 1996; Christiansen et al., 1998). One type of cue that is likely to be particularly useful in word segmentation is statistical information derived from the distribution of patterns of sounds (Hayes and Clark, 1970; Goodsitt et al., 1993; Christophe et al., 1994; Brent and Cartwright, 1996; Saffran et al., 1996a; Saffran et al., 1996b; Saffran et al., 1997; Aslin et al., 1998; Christiansen et al., 1998). The hypothesis that word boundaries might be discovered by computing the statistical properties of sound sequences in linguistic input has a long history, dating back to the eminent structural linguist Harris (1955). Harris noted that a number of different sounds can follow the last sound of a word. For example, the last sound in *elephant* might be followed by the first sound of any word that the grammar allows to occur next in the utterance. In contrast, the sequencing of sounds within a word is far more constrained. Given a sequence such as *ele*, there is a very strong expectation that the next sequence of sounds will be *phant* or *vator*.

This observation may be converted into a more precise statistic, *transitional probability* (Miller and Selfridge, 1950; Goodsitt et al., 1993; Saffran et al., 1996a). Along with other related statistics like conditional entropy, transitional probabilities track the contingencies between adjacent events: if event X occurs, what is the likelihood of event Y? This probability is computed by calculating the frequency with which X and Y co-occur, and then normalizing that frequency by the overall frequency of X. Returning to the domain of word segmentation, the transitional probability between two sounds, when tracked across a corpus of utterances, will generally be greatest when the two sounds are within the same word. When computed across word boundaries, probabilities of sound pairs should generally be lower, reflecting the decreased constraints at boundaries noted by Harris (1955). To the extent that these statistical cues are available, and they should be available in all

languages, given that words are characterized by internal coherence cross-linguistically, language learners equipped with the right computational tools should be able to use statistical information to detect word boundaries.

But are humans such learners? A wealth of statistical cues is of little use unless humans can detect and exploit them. In this case, learners must be able to compute statistical information in a fairly fine-grained way. To address this question, we conducted a series of studies that directly asked whether subjects can use statistical cues in the service of word segmentation. In our first experiment, adult subjects heard a synthesized speech stream generated from six multisyllabic nonsense words (such as *bupada* and *dutaba*, made from a set of 11 syllables) which were concatenated together in random order (Saffran et al., 1996b). Because the speech stream was synthesized and edited, it contained no cues to word boundaries except for the statistical properties inherent in the words, which distinguished word-internal sequences from the more accidental sequences spanning word boundaries. Following twenty-one minutes of exposure to this speech stream, adults demonstrated on a forced-choice test that they could distinguish sequences of sounds that were words from sequences of sounds made up of the same syllables but not forming words. A second set of studies demonstrated that first-grade children as well as adults were able to succeed on this task, even when the speech stream was presented incidentally while subjects were attending to another task (Saffran et al., 1997). These findings strongly suggest that humans possess statistical learning abilities which may play an important role in the discovery of word boundaries, and that this learning process proceeds automatically as a byproduct of mere exposure.

Our subsequent series of experiments asked whether the youngest language learners, and the ones for whom word segmentation is most crucial, can also detect statistical cues to word boundaries. We exposed 8-month-old infants to a speech stream generated by a smaller nonsense language containing four trisyllabic nonsense words (Saffran et al., 1996a). Given the attentional limitations of infant subjects, the speech stream used in this experiment was only two minutes long. As in the previous experiments, statistical information was the only available word boundary cue. Despite the brevity of this learning experience, eight-month-old infants successfully distinguished the familiar words from non-words (trisyllabic strings consisting of novel sequences of familiar syllables), and also from part-words (trisyllabic strings consisting of familiar but less statistically predictable sequences spanning a word boundary). An additional experiment confirmed that this learning process was based on the computation of the conditional probabilities of successive syllables, rather than a simpler computation of the frequencies of syllable co-occurrences (Aslin et al., 1998). All of these results strongly suggest that detection of sequential probabilities plays an important role in the process of word segmentation.

3. Tone segmentation?

Why might humans possess a learning mechanism which detects the boundaries between groups of elements based on sequential probabilities? One possibility is that

this statistical learning ability has evolved to subserve components of language acquisition. Certainly, the rapidity with which learners, particularly infants, were able to discern the statistical distributions of the speech streams in these experiments is consistent with the hypothesis that the mechanism underlying this computation is an adaptive specialization for language learning. Alternatively, perhaps this statistical learning mechanism is also triggered by input from other domains, and subserves segmentation processes in several different domains. If so, this mechanism should be able to perform similarly on learning tasks which are not language-based. Attunement to probabilistic patterns in the environment is widely observed across domains and across species (e.g. Hasher and Zacks, 1984; Gallistel, 1990; Kelly and Martin, 1994; Reber, 1993). Thus, the statistical learning mechanisms used by humans to process linguistic materials may also be used in the acquisition of some types of non-linguistic stimuli.

The present experiments consisted of a non-linguistic analogue of the word segmentation tasks which were summarized above. We created continuous tone streams by translating the nonsense words from our previous experiments into 'tone words' that had no phonetic content. We then exposed subjects to these streams, just as we had previously done for syllable streams. Crucially, the tone words were identical in their statistical structure to the syllable-based words used in our prior experiments, permitting direct comparisons between the acquisition of the same statistical distributions when presented as speech versus tones. To the extent that statistical structures implemented in different domains are learned similarly, we may conclude that a single segmentation mechanism, rather than a domain-specific mechanism, is at work.

4. Experiment 1

As discussed in the preceding section, an abundance of recent evidence supports the claim that humans can use the statistical properties of sound sequences to distinguish word-like units from other sequences in the speech stream. The present experiment asked whether human statistical learning abilities can also subserve the segmentation of non-linguistic stimuli. To address this question, we created a sound stream identical in its statistical properties to the syllable stream employed by Saffran et al., (1996b) (Experiment 1). To do this, we substituted a distinct tone for each of the 11 syllables from which our words were created (e.g. *bu* became the musical note D). Each of the six trisyllabic nonsense words (e.g. *bupada*) from the artificial speech language was thereby translated into a sequence of three musical notes (e.g. DFE). These 'tone words' were then concatenated together, in random order, to generate a continuous tone stream identical in statistical structure to the speech stream created by Saffran et al. (1996b).

A second tone stream was also generated and presented to a second set of subjects. This second tone stream (Language Two) contained the same 11 tones as the first set of tone words (Language One), but the tones were differently assigned to replace particular syllables. This created a different tone stream which had an overall sta-

tistical structure nearly identical to that of Language One. Subjects in both tone-language groups were then presented with the same test trials. Each test trial consisted of two three-tone test items, one of which was a tone word from Language One while the other was a tone word from Language Two. Sequences which were words for subjects exposed to Language One were non-words for subjects exposed to Language Two, and vice versa. The correct choice on each trial therefore depended upon whether the subject had been exposed to Language One or Language Two. This counterbalancing ensured that test performance across the two tone-language groups would reflect learning during the exposure period, rather than other biases favoring some tone sequences over others.

4.1. Method

4.1.1. Subjects

Twenty-four adult subjects with normal hearing participated in this experiment. In order to avoid possible effects of musical expertise on performance, all of the subjects were self-identified as non-musicians, and had not taken instrumental lessons, sung in choruses, or studied music theory since the seventh grade. Subjects were randomly assigned to the two different tone-language groups. Six additional subjects were tested but excluded from the analysis because their prior music expertise, as reported during the experimental session, exceeded this criterion. One additional subject was eliminated due to misunderstanding the test instructions. Subjects were paid $6 for their participation.

4.1.2. Materials

Tone sequences were constructed out of eleven pure tones of the same octave (starting at middle C within a chromatic set) and the same length (0.33 s), using the sine wave generator in SoundEdit 16. The tones were combined into groups of three to form six tone words (Language One: ADB, DFE, GG#A, FCF#, D#ED, CC#D). While some tones appeared in only one word, others occurred in multiple words. For example, D occurred in four different words, while G# occurred in only one word. The statistical structure of these words exactly mirrors that of the words used by Saffran et al. (1996a). The tone words were not constructed in accordance with the rules of standard musical composition, and did not resemble any paradigmatic melodic fragments (e.g., major and minor triads, or familiar three-tone sequences like the NBC television network's chimes).

The six tone words were concatenated together in random order, with no silent junctures between words, to create six different blocks containing 18 words each. A particular tone word was never produced twice in a row. The six blocks were in turn concatenated together to produce a seven minute continuous stream of tones. The tone sequence was tape-recorded directly from the sound output jack of a Quadra 650 computer. As in the linguistic materials used by Saffran et al. (1996b), there were no acoustic markers of word boundaries. An orthographic representation of the tone stream is analogous to the following: DFEFCF#CC#DD#EDGG#A. The only

consistent cue to the beginnings and ends of the tone words were the transitional probabilities between tones. Transitional probabilities between tones within words averaged 0.64 (range 0.25–1.00). In contrast, transitional probabilities between tones spanning word boundaries averaged 0.14 (range 0.05–0.60).[2]

The second tone language was constructed in precisely the same manner as the first. The same eleven tones were used, but combined differently to form six new words (Language Two: AC#E, F#G#E, GCD#, C#BA, C#FD, G#BA). The statistical structure of Language Two was very similar to Language One. Transitional probabilities between the tones within words averaged 0.71 (range 0.33–1.00), with lower average probabilities across word boundaries (mean = 0.18; range 0.07–0.53).[3]

To assess learning, we constructed a 36 item two-alternative forced-choice test exactly analogous to the test used by Saffran et al. (1996b). Each test item consisted of two tone-sequences: a word and a non-word. Non-words consisted of three-tone sequences also made of tones drawn from the language, but which had never occurred in that order during exposure (transitional probabilities = 0.0). One of the sequences presented on each trial was a word from Language One, and the other sequence was a word from Language Two. For a subject exposed to Language One, the non-words were words from Language Two; the opposite pattern obtained for subjects exposed to Language Two. If both languages were learned equally well, then we could be assured that performance on the test did not reflect perceptual biases that were unrelated to the statistical structure of the language. All six words from each of the two languages were paired exhaustively with one another, rendering 36 test trials. The two tone-sequences presented on each trial were separated by a 0.75 second pause, with an inter-trial interval of 5 s. In both conditions, two different random orders of the test trials were generated and each was used to test half of the subjects.

4.1.3. Apparatus

The study was conducted in a IAC sound-attenuated booth. The tone stream and the test were presented using an Aiwa tape deck and a Proton speaker.

4.2. Procedure

All subjects were run individually. Subjects were instructed that they would hear a

[2]These two distributions, transitional probabilities within words vs. across boundaries, overlap at the edge. This arises from the fact, as in real languages, that some syllable sequences appear both (regularly) inside of words and (occasionally) at word boundaries. However, in both of our tone languages this overlap was rare: in the present stimuli it occurred for only three of the 30 across-word tone-pairs. In one case, this probability was 0.6 (when the word GG#A happened to be followed by DFE), as the cross-boundary sequence AD also occurred in the word ADB. In the other two cases, this probability was 0.35, when either of the two tone words ending in D were followed by the word beginning with F, as the cross boundary sequence DF also occurred in the word DFE. Of course, despite this overlap, the mean transitional probability differed greatly between within-word and across-word tone-pairs. The occurrence of such coincidental overlaps makes segmentation more difficult, and therefore subjects' success at it even more notable.

[3]Only one of the probabilities spanning word boundaries, from A to C# (0.528), overlapped with the within-word probabilities, as this pair of tones also occurred inside the word AC#E.

tape of continuous tones. They were not told that the tape contained units of any sort. Subjects were asked to relax and to avoid consciously analyzing the tape while listening. However, they were also instructed not to entirely block out the sounds because they would be tested following the listening session. Subjects were not told which aspects of the tone sequence would be tested.

Subjects then listened to the seven-minute-long recording of one of the two tone-streams described above, repeated three times. Each of the three seven-minute listening sessions was followed by a short break. After listening for a total of 21 min, subjects received the forced-choice test. Subjects were instructed to indicate the most familiar tone sequence on each test-trial by circling either 1 or 2 on their answer sheet, corresponding to whether the familiar sequence was played first or second on that trial. The correct choice for subjects exposed to Language One was the incorrect choice for subjects exposed to Language Two. Half of the subjects exposed to each language received one randomized test order, while the other half received a different test order. All subjects heard three practice trials prior to testing.

4.3. Results and discussion

The overall results are presented in Fig. 1. As there were no significant differences between the two test orders for either Language One ($t(10) = 0.694$, n.s.) or Language Two ($t(10) = 1.87$, n.s.), data from the two test orders were pooled in the subsequent analyses. The mean score for subjects exposed to Language One was 26.5 out of a possible 36 (74%), where chance performance equals 18. A single

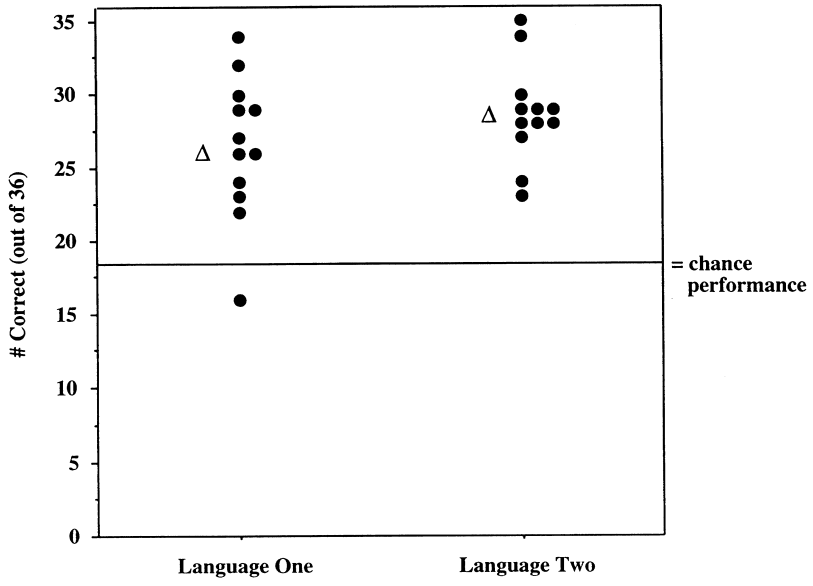

Fig. 1. Performance by adults on the two tone-languages in Experiment 1. Filled circles represent the number correct (out of a possible 36) for individual subjects in the word vs. non-word comparison. Open triangles represent the group means.

sample *t*-test (all tests two-tailed) revealed overall performance significantly different from chance: $t(11) = 6.05$, $P < 0.0001$. Five of the six words were learned at a level significantly better than chance ($P < 0.01$ for four words; $P < 0.05$ for one).[4] The mean score for subjects exposed to Language Two was 28.7 out of a possible 36 (80%). A single sample *t*-test revealed overall performance significantly different from chance: $t(11) = 10.8$, $P < 0.0001$. All six words were learned at a level significantly better than chance ($P < 0.01$ for five words; $P < 0.05$ for one). Although subjects performed slightly better on Language Two than on Language One, this difference was not significant: $t(22) = 1.26$, n.s. Since the two languages served as controls for one another (words from Language One were non-words for Language Two and vice versa), the lack of significant differences between language groups suggests that these results reflect learning of the statistical structure of the tone sequences presented during exposure.

Next, we asked whether the strengths of the transitional probabilities between tones played a role in determining how well particular words were learned (as observed by Saffran et al., 1996b). Recall that the transitional probabilities between pairs of tones within words ranged between 0.25 and 1.0. Since each word contained three tones, there were two transitional probabilities associated with each word: between the first and second tones and between the second and third tones. For the following analysis, the average of the two transitional probabilities for each word was computed. We split the six words of each language into two sets, one set containing the three words with the highest average transitional probabilities (1.0, 0.75, and 0.75 for both languages), and the other set containing the three words with the lowest average transitional probabilities (0.425, 0.425, and 0.5 for Language One; 0.43, 0.67, and 0.67 for Language Two). For each language, an ANOVA comparing subjects' mean scores on the three high probability words with the three low probability words was performed. As with the linguistic materials used by Saffran et al. (1996a), performance was significantly better on the words of each language containing higher transitional probabilities: Language One: $F(5,11) = 2.67$, $P < 0.05$; Language Two: $F(5,11) = 7.31$, $P < 0.0001$.

Since the tone languages in this study were designed to be analogous to the artificial speech language used by Saffran et al. (1996b), we compared the present results with the non-word condition results from Saffran et al. (1996b) (Experiment 1). Fig. 2 illustrates the similarity of the results of these two experiments across domains. With linguistic materials (Saffran et al., 1996b), subjects scored an average of 27.2 out of a possible 36 (76%). The overall mean score for subjects in the present study was 27.6 out of a possible 36 (77%). A *t*-test comparing the results of the present study with the non-word results from Saffran et al. (1996b) revealed no significant differences between the speech stimuli and the tone stimuli: $t(34) = 0.26$, n.s.

An additional analysis examined the correlation between scores for each word

[4]Interestingly, the one tone word that was not learned was the same one (ADB) which contained a sequence that also occurred spanning a word boundary, as discussed in footnote 2. However, other tone words with this overlap feature *were* learned significantly above chance.

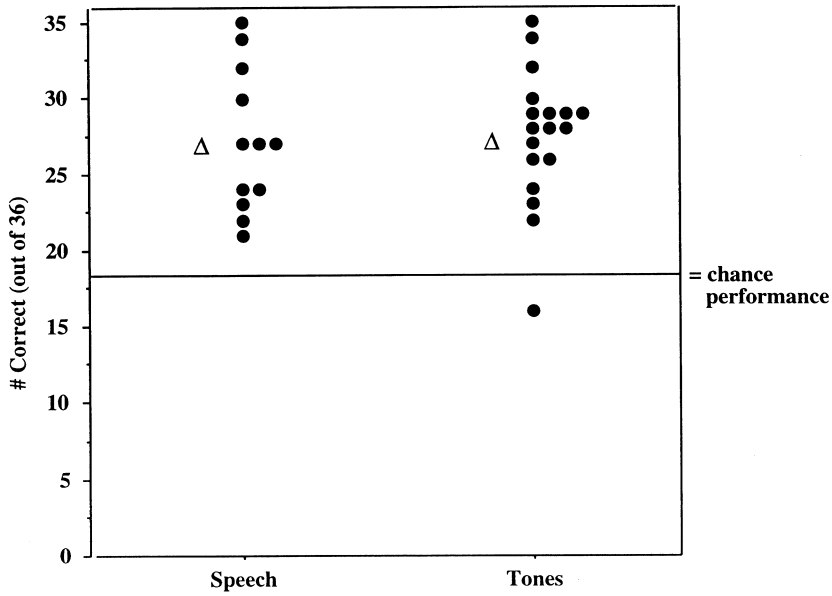

Fig. 2. Performance by adults on the word vs. non-word comparison embedded within either a speech stream (data from Saffran et al., 1996a) or a tone stream. Filled circles represent the number correct (out of a possible 36) for individual. Open triangles represent the group means.

when presented as speech (Saffran et al., 1996b; Experiment 1, non-word condition) and as tones (Language One in the present experiment). Because there were only six comparisons, this analysis lacked power; nevertheless, there was a strong trend towards a significant correlation ($R = 0.74$, R-squared = 0.54, $F(1,4) = 4.74$, $P < 0.10$). The individual words were learned similarly across the two domains, suggesting that the particular mode of presentation did not affect which words (or statistical structures) were learned better than others. This comparison supports the hypothesis that the same learning mechanisms underlie statistical learning of patterned stimuli across these two domains.

These results suggest that adult learners readily group sequences of auditory events in the same manner, regardless of whether the input is linguistic (syllables) or non-linguistic (tones). The finding that words with higher transitional probabilities are learned best, whether instantiated in tones or syllables, supports the hypothesis that a mechanism computing such probabilistic information is implicated in this segmentation process. The next experiment further explored the parallel between linguistic and non-linguistic learning by asking subjects to perform a more difficult discrimination following learning: distinguishing tone words from tone sequences containing parts of words. Using speech stimuli, Saffran et al. (1996b); Experiment 2) found that adults could distinguish words from part-words following exposure to a synthetic speech stream. We hypothesized that if adults can apply the same learning mechanism to continuous tone streams, then they should also be able to perform this word/part-word discrimination.

5. Experiment 2

5.1. Method

5.1.1. Subjects

Twenty-four adult subjects with normal hearing participated in this experiment. One additional subject was excluded due to equipment error. In order to avoid possible effects of musical expertise on performance, subjects were self-identified as non-musicians and had not taken instrumental lessons, sung in choruses, or studied music theory since the seventh grade. Subjects were randomly assigned to the two different exposure conditions, and were paid $6 for their participation.

5.1.2. Materials

Language One from Experiment 1 served as Language One in the present experiment. A second language was also constructed, using the same set of 11 tones. This language was designed to render a set of test items pitting words against *part-words*. A part-word, on analogy with Saffran et al. (1996b), consisted of a three-tone sequence comprised of two tones from a word plus a third tone. For example, consider the word ADB. To generate a part-word, either the first or third tone was substituted with a different tone (e.g. G#DB). Three part-words contained the first two tones of words plus a new third tone, and three contained the final two tones of words plus a new first tone. The transitional probabilities between the new tone and the two tones taken from a word were always zero.

Language Two was designed so that its words were the part-words with respect to Language One, while words from Language One were part-words with respect to Language Two. For example, consider the following pair of tone words: ADB and G#DB. For learners of Language One, where ADB was a word, G#DB was a part-word (with G# substituted for A). However, for learners of Language Two, G#DB was a word and ADB was a part-word. The tone words for Language Two were G#DB, DFF#, FG#A, C#CF#, D#EG#, and CC#B. The resulting statistical structure of Language Two was very similar to that of Language One. Transitional probabilities between tones within words averaged 0.56 (range 0.33–1.00) versus 0.64 (range 0.25–1.00) for Language One, and transitional probabilities between tones that spanned a word boundary averaged 0.15 (range: 0.067–0.40) versus 0.14 (range 0.05–0.60). As in Experiment 1, test sequences consisted of a pair of words, one from each language. The correct choices on the test thus depended upon whether the subject had been exposed to Language One or Language Two.

The test phase was conducted as in Experiment 1, but here the six words from Language One were paired exhaustively with the six words from the new Language Two to form the 36 test trials. Subjects were required to discriminate words from part-words, a more difficult task than the word vs. non-word distinction tested in the previous experiment, due to the greater similarity between words and part-words.

5.1.3. Apparatus and procedure

The apparatus and procedure were identical to Experiment 1.

5.2. Results and discussion

The overall results are presented in Fig. 3. As there were no significant differences between the two randomized test orders for either Language One ($t(10) = 0.38$, n.s.) or Language Two ($t(10) = 0.80$, n.s.), data from the two test orders were pooled in the subsequent analyses. The mean score for subjects exposed to Language One was 23.0 out of a possible 36 (64%), where chance performance equals 18. A single sample *t*-test (all tests two-tailed) revealed overall performance significantly different from chance: $t(11) = 3.91$, $P < 0.01$. Four of the six words were learned at a level significantly better than chance ($P < 0.01$ for one word; $P < 0.05$ for three words). The mean score for subjects exposed to Language Two was 23.7 out of a possible 36 (65.8%). A single sample t-test revealed overall performance significantly different from chance: $t(11) = 5.16$, $P < 0.001$. Four of the six words were learned at a level significantly better than chance ($P < 0.01$ for two words; $P < 0.05$ for two words). Although overall performance was slightly better on Language Two than on Language One, this difference was not significant: $t(22) = 0.44$, n.s. Consistent with the results of Saffran et al. (1996a), subjects tested on part-words did not perform as well as the subjects tested on non-words in Experiment 1: $t(46) = 3.49$, $P < 0.01$, presumably due to the greater difficulty of the part-word test.

Since the structure of the test stimuli in this study was exactly analogous to the part-word test from Saffran et al. (1996b) (Experiment 2) using speech stimuli, we

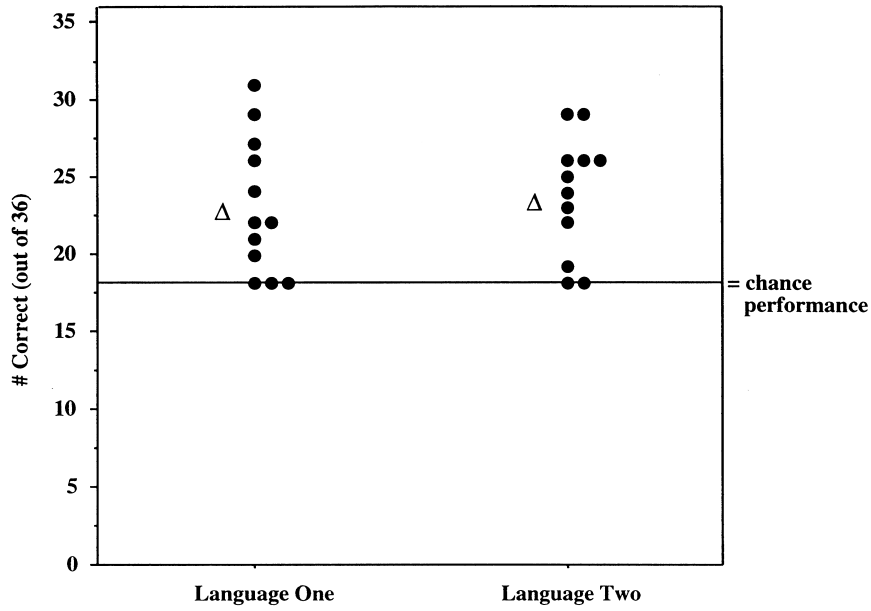

Fig. 3. Performance by adults on the two tone-languages in Experiment 2. Filled circles represent the number correct (out of a possible 36) for individual subjects in the word vs. part-word comparison. Open triangles represent the group means.

compared the results of these two experiments (see Fig. 4). With linguistic stimuli (Saffran et al., 1996b), subjects scored an average of 22.3 out of a possible 36 (62%) on the part-word test. The overall mean score for subjects in the present study was 23.4 out of a possible 36 (65%). A t-test comparing the results of the present study with the part-word results of Saffran et al. (1996b) revealed that the total scores for speech versus tones were not significantly different: $t(34) = 0.76$, n.s.

As with the results of Experiment 1, an additional analysis examined the correlation between total scores for each word when presented as speech (Saffran et al., 1996b) and as tones (Language One in the present experiment). Despite the low power due to the presence of only six words, there was a significant correlation between scores on words presented as speech and as tones: $R = 0.939$, R-squared = 0.88, $F(1,4) = 29.75$, $P < 0.01$. This correlation is striking, as it suggests that the ease with which particular sequences are learned depends on the statistical structure of the sequences, rather than the domain within which they are exemplified.

We performed one final comparison with the results of Saffran et al. (1996b) by examining the patterns of false-alarms to particular words. Subjects are most likely to false-alarm, and incorrectly choose the part-word over the word, when the part-word is perceived as highly similar to the material learned during exposure. Saffran et al. (1996b) found that subjects were most likely to incorrectly select those part-words which contained the final two syllables of a word rather than the initial two syllables of a word. An examination of the present data found that the part-words consisting of the final syllables of tone words were incorrectly chosen more fre-

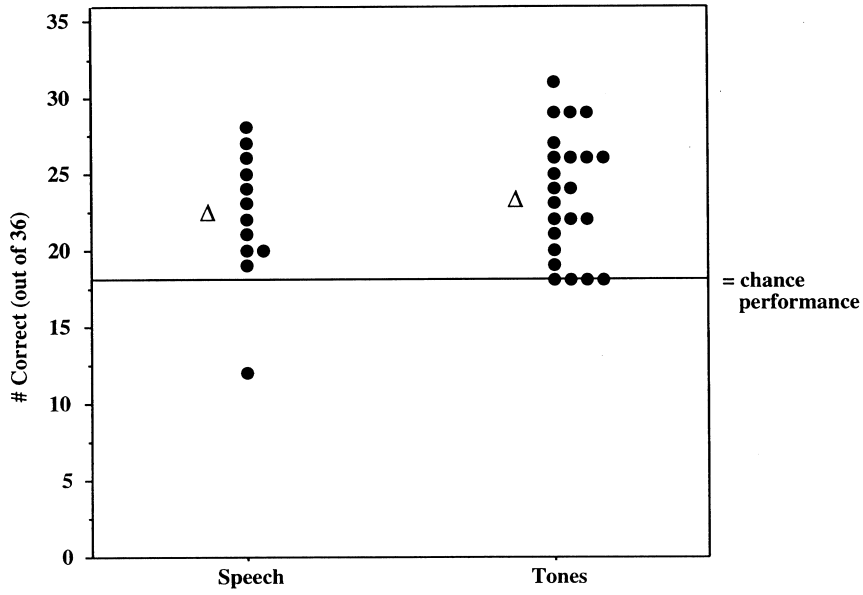

Fig. 4. Performance by adults on the word vs. part-word comparison embedded within either a speech stream (data from Saffran et al., 1996a) or a tone stream. Filled circles represent the number correct (out of a possible 36). Open triangles represent the group means.

quently than those consisting of the initial syllables of words: $t(23) = 2.28$, $P < 0.05$. As with the linguistic stimuli, subjects were more likely to confuse words with part-words when the ends of words were identical, suggesting that learners may acquire the ends of words first, even when they consist of tone sequences.

While these analyses strongly suggest that these tone stimuli were learned in a manner analogous to the speech stimuli, there are some differences between linguistic and musical stimuli which merit exploration. In particular, sequences of tones contain harmonic structure in the form of musical intervals or relative pitch relations. A melodic sequence like ADB contains not only these three pitches but also two intervals: a descending perfect fifth between A and D, and an ascending perfect sixth between D and B. Such intervals are relative in nature: ADB and BEC# contain the same pair of intervals, despite their different absolute pitches. Corresponding linguistic stimuli, such as those used by Saffran et al. (1996b), do not transparently contain any analogous relations[5]. We therefore analyzed the melodic structure of the tone words used in the present experiment in order to determine whether test performance reflects the use of the harmonic relations (intervals) in these stimuli, rather than the probabilities with which different pitches followed one another (analogous to the linguistic version of the task). While the tone stimuli were explicitly designed not to conform to paradigmatic melodic fragments (e.g. major and minor triads, or tone sequences mirroring commonly known fragments like the NBC television network's three tone chimes), it is possible that other features of the melodic structure of these stimuli could have accounted for above-chance performance on the part-word test.

One possibility is that subjects were able to distinguish tone words from tone part-words because the part-word sequences contained different intervals than the word sequences. Table 1 lists the distribution of the intervals from the words of Languages One and Two. These distributions are similar, with ascending minor seconds occurring most frequently in both languages. In order to further explore the possible role for different intervalic contents of words and part-words, we examined the false-alarm rates for part-words which contained intervals not present in the words. If performance on the test reflected the distinction between intervals which occurred within words versus intervals which did not, one would expect that subjects should be least likely to false-alarm when the part-word contained novel intervals[6], while false-alarms should have occurred most often for part-words containing the same intervals as words. This pattern was not found in the data: false-alarm rates to part-words containing novel intervals ranged from the highest (2.83/6) to the lowest (1.25/6) across all of the part-words from the two languages, regardless of their interval structure. These results suggest that performance was not a simple function

[5]It is possible that syllables are structurally related in such a way that some syllable-pairs are more naturally similar or perceptually distinct than other syllable pairs. However, no such metric has yet been described for adults.

[6]Note that for the purposes of this discussion, a novel interval is considered to be an interval which did not occur word-internally. Other intervals did occur across word-boundaries; however, the frequency of these intervals was lower, and they are not included in the present analysis.

of noting which intervals occurred or did not occur in the corpus of tones vs. the test sequences.

Another related tone-interval hypothesis is that subjects may have performed the task by tracking which intervals followed one another. Rather than construing tone words as short three-note melodies, they might be perceived as two-interval sequences. The next set of analyses asked whether subjects kept track of which intervals followed one another. On this account, words that contained more frequent interval pairs should have been learned best. In each language, two of the words contained the same pair of intervals, making this interval sequence the most frequent in the language (Language One: two ascending minor seconds; Language Two, an ascending minor third followed by an ascending minor second). If the most frequent interval sequences were learned best by subjects, then these two words in each language should show the best performance. Although ANOVAs demonstrated that there were no significant differences in mean scores across words for each language (Language One: $F(5,11) = 1.19$, n.s.; Language Two: $F(5,11) = 1.84$, n.s.), we examined the ordinal rankings of mean scores for these word pairs in each language to see if words sharing the same sequence of intervals were learned best. In both languages, one of the two words which shared the same interval pair was learned best (Language One: CC#D, M = 4.33; Language Two: FG#A, M = 4.50). However, the scores for the other word which shared the same interval pair fell into the middle of the distribution of scores (Language One: GG#A, M = 4.08; Language Two: DFF#, M = 4.08). This pattern of results suggests that subjects did not treat sequences of identical intervals differing in absolute pitch equivalently, although we cannot rule out the possibility that interval information contributed to the tone segmentation process.

As in Experiment 1, the results of this second experiment strongly support the hypothesis that the same learning processes underlie the grouping of sequences of syllables and tones. In particular, the statistical structure of the words appears to account for the learning process across both of these domains. The similarity of the findings for tones and syllables, including even which words (in terms of their statistical structure) were learned best, points to the conclusion that the acquisition of auditory sequences is strongly influenced by their statistical structure, regardless of the domain within which they are presented (see also Altmann et al., 1995).

One surprising feature of these results is the finding that part-words which resembled the ends of words were more likely to be confused with words than part-words resembling the beginnings of words. This result replicates a similar finding with speech stimuli (Saffran et al., 1996b), and suggests that the ends of words are learned first, whether the words are created from syllables or tones. A large body of literature focusing on word learning under more naturalistic circumstances also suggests that the ends of words are privileged during acquisition (e.g. Slobin, 1973; Echols and Newport, 1992; Golinkoff and Alioto, 1995). While the most obvious explanation is that the ends of words are particularly salient because they are followed by silence at utterance boundaries, our results using continuous speech and tone streams cannot be accounted for in this fashion, as no cues from silence were available. Saffran et al. (1996b) hypothesized that the end-of-word

superiority effect might be a side effect of the use of learning mechanisms which compute forward transitional probabilities: the final syllable anchors the transitional probability computation that discovers word boundaries, as its frequency serves as the denominator for this computation.[7] On this view, the ends of words may be privileged with respect to the beginnings of words because learners may be most likely to maintain the end-of-word information that anchored the pertinent computation. The finding that non-linguistic stimuli elicit this same pattern of results provides additional evidence for probability-based learning across these two domains.

In our final study, we asked how these results from adult subjects compare to the mechanisms available to infants learners, for whom word segmentation is a critical component of native language acquisition. It is possible that the present results reflect processes unrelated to child language acquisition, whereby adults can strategically apply domain-specific mechanisms to other learning problems when necessary. Alternatively, the application of learning mechanisms to multiple domains in adulthood might reflect an analogous ability available during development. In the latter case, domain-specific knowledge reflected in many fluent processes in adults might arise, at least in part, from mechanisms that were not tailored to solve domain-specific problems. To address these issues, we asked whether infants, like adults, could segment non-linguistic sound sequences by applying statistical learning mechanisms known to be used in the segmentation of linguistic stimuli (Saffran et al., 1996a; Aslin et al., 1998).

6. Experiment 3

In this study, we exposed infant subjects to continuous tone streams, created by translating speech streams used successfully in prior infant segmentation experiments (Saffran et al., 1996a) into tones. Infants were first familiarized with a tone stream, which served as a brief learning experience (Jusczyk and Aslin, 1995). Learning was then assessed using the preferential listening methodology (e.g. Kemler Nelson et al., 1995). The tone streams used in this experiment were identical in their statistical structure to the speech streams designed by Saffran et al. (1996a) (Experiment 2), with a tone substituted for each syllable. Following familiarization, we assessed infant listening preferences for tone words versus part-words, sequences of tones spanning word boundaries. If infants did not learn the statistical patterns of tones heard during familiarization, then no differences in listening times for words versus part-words should emerge during testing. If, however, infants were able to track the statistical co-occurrence of tones, then a difference in listening times for the more familiar words versus part-words might be expected, as observed previously with linguistic stimuli (Saffran et al., 1996a). Specifically, we expected longer listening times to the novel part-words than to the familiar words, a novelty

[7] We have only considered the computation of forward transitional probabilities (e.g. Y|X/X). If infants compute backward transitional probabilities (e.g. X|Y/Y), then low values will occur at the first syllable of a word rather than at the last syllable of a word.

Table 1
The numbers in each column represent how many words (out of a possible six) contained each musical interval.

Interval	Language One	Language Two
Ascending		
Minor second	5	4
Minor third	1	2
Major third	0	1
Augmented fourth	1	1
Major sixth	1	1
Minor seventh	0	1
Descending		
Minor second	1	1
Major second	1	0
Perfect fourth	1	0
Augmented fourth	0	1
Perfect fifth	1	0

effect that we have observed in all of our previous infant experiments using speech stimuli.

6.1. Method

6.1.1. Subjects

Two groups of twelve 8-month-old infants were tested (mean age 7 months 4 weeks; range 7:1 to 8:2). An additional 18 infants did not complete the experiment due to fussiness. All infants were solicited from local birth announcements and hospital records, and parental consent was obtained prior to testing in accordance with the guidelines of the local human subjects review committee and the principles of ethical treatment established by the American Psychological Association.

6.1.2. Stimuli

As in the previous two experiments, two tone streams were created, with each tone.33 sec in duration. Each language consisted of four tone words (Language One: AFB, F#A#D, EGD#, CG#C#; Language Two: D#CG#, C#EG, FBF#, A#DA). For each language, 45 tokens of each tone word were concatenated together in random order to create a 3-minute tone stream, with the stipulation that the same tone word never occurred twice in a row. Because each tone was longer in duration than the syllabic stimuli used by Saffran et al. (1996a), which were presented at a rate of 4.5 syllables per second, the 180 tone words took one minute longer to play than the 2-min stream of 180 syllabic words used in this previous speech experiment. The stimuli were created using the tone generator in SoundEdit 16, and digitized at a sampling rate of 22 kHz for on-line playback through an Audiomedia sound-board in a Quadra 650 computer.

Testing was performed not with a 2-alternative forced-choice task (as in adults), but rather by presenting a single test item (repeatedly) on each test trial, and then

comparing the infants' responses to the two different types of items over a series of test trials. Each test item consisted of a three-tone sequence. The same four test items were used for all infants (AFB, F#A#D, D#CG#, C#EG). Two of these test items were tone words from the familiarization language, while the other two were tone part-words. In this experiment, as in the infant speech study by Saffran et al. (1996a); Experiment 2), a part-word consisted of a three-tone sequence spanning a word boundary. Part-words were created by joining the final tone of one word to the first two tones of another word. Thus the part-word sequences were heard during familiarization. However, their statistical properties differed from the words. Specifically, tone-pairs within words had transitional probabilities of 1.00 and 1.00, whereas tone-pairs within part-words had transitional probabilities of 0.33 and 1.00. In addition, each word was presented 45 times in the familiarization corpus, whereas the random ordering of words resulted in 15 instances of each part-word. Testing thus asks whether infants can discriminate tone words from tone part-words on these statistical bases. For infants exposed to Language One, AFB and F#A#D were words and D#CG# and C#EG were part-words, with the opposite pattern for infants exposed to Language Two. This between-subjects counterbalanced design ensured than any observed preferences for words or part-words across the two languages resulted from statistical learning, and not from any inherent preferences for certain tone sequences.

6.1.3. Procedure

Each infant was tested individually while seated in a parent's lap in a sound-attenuated booth. An observer outside the booth monitored the infant's looking behavior on a closed-circuit TV system and coded the infant's behavior using a button-box connected to the computer. This button-box was used to initiate trials and to enter the direction of the infant's head turns, which controlled the duration of each test trial. Both the parent and the observer listened to masking music over headphones to eliminate bias. Infants were randomly assigned to hear either Language 1 or Language 2. At the beginning of the 3-minute familiarization phase, the infant's gaze was first directed to a blinking light on the front wall in the testing booth. Then the sound sequence for one of the two tone languages was presented without interruption from two loudspeakers (one located on each of the two side walls in the booth). During this familiarization period, to keep the infants' interest, a blinking light above one of the two loudspeakers (randomly selected) was lit and extinguished dependent on the infant's looking behavior. When this blinking side-light was extinguished, the central blinking light was illuminated until the infant's gaze returned to center, and another blinking side light was presented to elicit the infant's gaze. During this entire familiarization phase there was no contingency between lights and sound, which played continuously.[8] Immediately after familiarization, 12

[8]This familiarization procedure differs from that used by Jusczyk and Aslin (1995) in that the presentation of the blinking lights, but not the auditory stimuli, was contingent upon the infants' looking behavior. Thus, during familiarization the infants may have learned the contingency between their looking behavior and the presentation of the blinking lights, although their looking behavior was not related to the presentation of the auditory stimuli. The sound stream presented during familiarization may be viewed as an incidental background task, analogous to procedures used with young children and adults (Saffran et al., 1997).

test trials were presented (three trials for each of the four test items, presented in random order). Six of these trials were tone words and six were tone part-words. Each test trial began with the blinking light on the front wall. When the observer signaled the computer that the infant was fixating this central light, one of the lights on the two side walls began to blink and the central light was extinguished. When the observer judged that the infant had made a head turn of at least 30 deg in the direction of the blinking side light, a button press signaled to the computer that one of the test items should be presented from the loudspeaker adjacent to the blinking light. This test item was repeated with a 500 ms interstimulus interval until the observer coded the infant's head turn as deviating away from the blinking light for 2 consecutive s. When this look-away criterion was met, the computer extinguished the blinking side light, turned off the test stimulus, and turned on the central blinking light to begin another test trial. The computer randomized the order of test trials (three for each of the four test items) and accumulated total looking time to each of the two test words and two part-words.

6.2. Results and discussion

Looking times for words and part-words were averaged across the two language groups because no differences were observed between Languages One and Two: $t(22) = 1.02$, n.s. As in Saffran et al. (1996a), infants showed a significant difference in listening times to the two types of test items (part-words versus words): $t(23) = 2.14$, $P < 0.05$. Listening times to part-words (M = 6.92, SE = 0.48) exceeded listening times to words (M = 5.88, SE = 0.45), the same pattern found in the corresponding study using speech stimuli by Saffran et al. (1996a). This difference demonstrates that infants are able to distinguish sequences which form words from sequences which span word boundaries, even when those sequences are presented as tones rather than syllables.

An ANOVA was conducted to compare the performance of these infants tested with tone sequences to the infants from Saffran et al. (1996a) (Experiment 2) tested with syllable sequences. There was no main effect of tone vs. speech domain ($F(1,46) = 1.9$, $P = 0.17$, n.s.), but there was a significant main effect of test item (word vs. part-word) ($F(1,46) = 9.87$, $P < 0.01$). Finally, there was no interaction between domain (tone vs. speech) and test item (word vs. part-word) ($F(1,46) = 0.13$, $P =$ n.s).

The basis for infants' ability to discriminate words from part-words in tone sequences, therefore, is likely to reside in the same sorts of probability computations between speech sounds discussed by Saffran and colleagues (Saffran et al., 1996a,b, 1997), and directly tested by Aslin et al. (1998). That is, infants are presumably computing the transitional probabilities between adjacent tones in the tone stream, grouping tones with high transitional probabilities, and forming tone-boundaries at locations in the tone stream where transitional probabilities are low. A related type of information that infants might be using to distinguish words from part-words is the statistical properties of patterns of intervals within and between words. Although our previous analyses suggest that adults do not rely on such tone-interval informa-

tion, the adult tone-languages were sufficiently different from the infant tone-languages that this tone-interval explanation cannot be ruled out for infants. To use interval information in this fashion, infants would be required to distinguish the most frequent interval pairings (words) from interval pairings which occurred less often (part-words). For example, consider the test item D#CG#, an interval sequence comprised of a descending minor third followed by an ascending minor sixth. This sequence was a word for infants exposed to Language Two and a part-word for infants exposed to Language One. For infants exposed to Language One, this part-word occurred during familiarization when the word CG#C# was preceded by the word EGD#, as it was for one-third of its occurrences. The minor sixth derived from the word CG#C occurred 45 times during the familiarization session, while the minor third derived from the concatenation of EDG# and CG#C# occurred 15 times during the familiarization session. Thus, given infants' sensitivity to musical intervals (Schellenberg and Trehub, 1996), if they kept track of the frequencies of interval sequences, the distinction between words and part-words might have emerged (note, however, that preferences for specific intervals could not account for test performance because of the counterbalancing of words and part-words in the two tone languages). While we suspect that infants were in fact computing transitional probabilities of particular tone sequences rather than interval sequences, the present data cannot distinguish these two possibilities. Either way, infants are evidently able to bring their statistical learning abilities to bear on tonal sequences.

7. General discussion

These three experiments offer striking evidence for the similarity of statistical learning in segmenting tone sequences and syllable sequences. Adult subjects in the first two experiments showed levels of performance on a tone segmentation task equivalent to subjects in the analogous speech segmentation task studied by Saffran et al. (1996b). The parallels in performance reached beyond overall scores, including, most notably, the finding that the statistical structure of particular sound sequences dictated the outcome of learning. There were no differences in performance attributable to the particular domain within which the statistical learning task was implemented. Importantly, this domain-independence was also found with eight-month-old subjects, whose results on the tone segmentation task paralleled the results from the analogous speech segmentation task studied by Saffran et al. (1996a). These findings suggest that linguistic stimuli are not privileged with respect to tones as input to this particular statistical learning process, despite the fact that spoken word segmentation is integral to language acquisition, while the significance of tone sequence segmentation is less evident.

Before we can conclude that the mechanism previously identified as available for linguistic segmentation is also available for tone segmentation, two obvious questions must be addressed. The first pertains to the linguistic task to which we compared the present non-linguistic task. Was the original task designed by Saffran et al.

(1996a,b) actually a study of language learning? By virtue of human anatomy, speech from natural languages is never continuously spoken for 21 min, or even 2 min, without breaths or other pauses. Similarly, human speech is not characterized by the lack of prosodic and phonological variability found in our synthetic speech streams, where all of the vowels were full and the pitch was invariant. One might therefore suggest that the findings purportedly pertaining to linguistic stimuli were in fact not based on linguistic processing. If this is the case, then the similar learning outcomes for the synthetic speech and tones might reflect the possibility that both utilize a non-linguistic learning mechanism, rather than the same mechanism for linguistic and non-linguistic inputs. However, the subjective experiences of adult participants in the speech segmentation experiments strongly suggest that these stimuli were perceived as linguistic: when asked, subjects were able to transcribe the speech orthographically, and wondered how the experimenters were able to remove breaths from the speech. We have no doubt that subjects asked to listen to the speech stimuli and the tone stimuli would not hesitate in labeling the former as language and the latter as music.

The second possibility is that because some languages use tone contrastively, perhaps the tone sequences were perceived as linguistic.[9] If this were the case, then the present tone experiments were not in fact a test of non-linguistic statistical learning. However, our tone stimuli, unlike input derived from tone languages like Mandarin, contained no phonetic content. Moreover, even when phonetic content is available while subjects are engaged in *pitch* processing tasks, right hemisphere activation, as opposed to the left hemisphere activation resulting from *phonetic* processing tasks using the same stimuli, is present (Blumstein and Cooper, 1974; Zatorre et al., 1992). Given these considerations, it seems quite unlikely that subjects processed the tone streams as linguistic materials. However, we recognize that our results do not definitively eliminate the possibility that the processing of tone sequences is captured by a language-specific mechanism.

In short, our results suggest that the same statistical learning mechanism can operate on both linguistic and non-linguistic stimuli. Thus at least part of the machinery involved in natural language learning may be shared with other pattern learning processes. However, we do not yet know how these results may generalize to the larger question of domain-specificity or domain-generality for language acquisition as a whole. The present results entail several important restrictions. First, the studies reported here compare the processing of speech streams to that of tone streams. The similarity of outcomes may thus pertain only to the segmentation of sequential auditory materials, and may not be entirely domain-general. We do not yet know, for example, whether the same type of statistical learning process operates across modalities, or across different types of patterns (for example, materials in which the patterning occurs across space rather than through temporal sequencing). Other literature has suggested similarities between language and music which may be particular to these two highly structured, and specifically human, arenas (Lerdahl and Jackendoff, 1983; Jusczyk and Krumhansl, 1993; Tre-

[9]We thank Peter Gordon for this suggestion.

hub and Trainor, 1993). However, ongoing work in our labs does suggest that our findings are not limited to auditory materials, and show many of the same properties for visual pattern learning (Asaad, 1998) and visuo-motor sequence learning (Hunt and Aslin, 1998).

Second, and perhaps most important for future research, our results thus far focus on the process of segmentation. The problem of segmenting elementary units out of a large and apparently unsegmented input is a very general problem, characteristic not only of early levels of language processing, but also of perception in many domains. It is thus possible that the process of segmentation may have similar solutions across these domains (e.g. 'Group together items which tend to co-occur, and segment at points where predictability declines'), and for this reason be handled by common (or separate but analogously functioning) mechanisms. At the same time, language acquisition includes many other problems, in addition to that of segmentation; for example, forming grammatical categories and acquiring hierarchical phrase structure. These additional parts of language acquisition clearly require mechanisms other than those that compute transitional probabilities. As Chomsky noted many years ago (Chomsky, 1957), in contrast to word segmentation, other aspects of natural language structure cannot be described by finite state regularities and therefore cannot be acquired by a mechanism limited to the computation of transitional probabilities. Mechanisms for learning these other aspects of linguistic structure must be capable of computing quite different types of statistical regularities (Maratsos and Chalkley, 1980; Morgan and Newport, 1981; Morgan et al., 1987; Mintz et al., 1995; Cartwright and Brent, 1997; Saffran, 1997), or perhaps they are quite different types of mechanisms altogether (Marcus, 1998). Such mechanisms may or may not be shared with the processing of other domains. Our findings suggesting a common statistical learning mechanism for speech and tone segmentation therefore do not imply either that higher levels of language are acquired by the same mechanisms which perform segmentation, or that these mechanisms are domain-general. Along with other investigators, we are engaged in on-going research investigating the types of statistical learning procedures which might be employed for such tasks (Mintz et al., 1995; Saffran, submitted), but much future research will be required before one can address the domain specificity issue more generally.

A final caution concerns the relation of the present results to questions of innate constraints on learning, and also to the question of how to explain the acquisition of domain-specific bodies of knowledge from (at least partly) domain-general beginnings. It is clear from even a cursory examination of linguistic theory, music theory, and ecological optics that the structure of knowledge ultimately achieved in these domains is distinct: Language, music, vision, and the like entail distinct primitives (e.g. phonemes, syllables, and phrases, versus pitches and intervals), as well as apparently distinct combinatorial principles (cf. the principles of morphology and syntax as compared with the principles of harmonic and melodic structure). How can one explain these differences in knowledge? The traditional account has been to claim that virtually all of the processing and learning of these domains is handled by distinct mechanisms, which are thought to be innately specialized for handling the

tasks of their particular domain (Fodor et al., 1974, 1983; Chomsky, 1975b). The usual alternative account has been an anti-nativist one, which expects that nearly all knowledge can be acquired by a common set of processes and mechanisms (Elman et al., 1996; Seidenberg, 1997). As we have noted, the present results do not directly support or conflict with either of these larger positions, since they focus only on one piece of the domains, namely, initial segmentation. Nonetheless, it may be worth mentioning briefly how our results might fit into a larger position. One possibility is that a common set of initial segmentation processes, applied to distinct perceptual inputs, feeds into subsequent mechanisms which are entirely distinct. A second possibility is that all of the processes for language and music are shared, and that the apparently different organizational structure of languages and musical systems is entirely the accidental result of the different regularities of the perceptual inputs in these domains. Our own view, however, is toward a third possibility: Learning mechanisms, by virtue of their architectures, always compute and acquire certain kinds of patterns more readily than others. In this sense, there are *always* innate constraints on learning; no learning mechanism, however powerful, learns every type of pattern equally well. In the present case, our results suggest that human infants and adults possess a mechanism which readily and rapidly computes transitional probabilities among sequences of *adjacent* units, regardless of whether the units themselves are syllables or tones; but ongoing work suggests that this mechanism is quite selective in the types of *non-adjacent* regularities it can compute (Calandra, 1998; Newport and Aslin, in press). Similarly, as researchers begin to achieve a better understanding of the computational machinery underlying other aspects of language and music acquisition, we presume that they will discover selectivities of what can be computed and remembered. For example, language and music may differ on dimensions such as the sequential versus simultaneous nature of their elements, or the continuous versus discrete nature of element modulations, thereby producing different organizational patterns in the languages and musical genres which are readily acquired and retained. Modest computational biases in the machinery that learns in these domains may therefore play an important role in distinguishing the domains. Whether this type of account will be adequate, or which of the three accounts we have mentioned will apply to various aspects of knowledge acquisition, must await future research.

In summary, our findings support the hypothesis that learning of sequential dependencies in auditory stimuli involves a mechanism that can be deployed to group and segment both speech and tone elements. Further research in other domains (e.g. vision and visuomotor) will clarify the extent to which the statistical learning observed in segmentation in the auditory modality is best characterized as a domain-general mechanism. Of course, we do not claim that a domain-general learning mechanism is sufficient to account for all levels of language processing. It seems likely that a variety of highly constrained learning mechanisms, at least some of which are specific to humans and to language, will be needed to account for language processing and acquisition as a whole. The segmentation of words from fluent speech, and tones from tone sequences, is a relatively low-level aspect of statistical learning which forms one of the first steps in analyzing these patterned

domains. More complex and higher level aspects of language and musical structure will require comparably more complex (and potentially quite different) types of computations to learn them. Nonetheless, we believe that our experiments on statistical learning of word and tone segmentation may make a new contribution to describing the learning mechanisms and the range of constraints required to learn various aspects of language.

Acknowledgements

We thank J. Gallipeau and J. Hooker for their help in testing the infants. Portions of Experiments 1 and 2 were based on a senior honors thesis by Elizabeth K. Johnson. Support was provided by a National Science Foundation predoctoral fellowship to JRS, a NSF REU grant to EKJ, a NSF research grant to RNA (SBR-9421064), and a National Institutes of Health research grant to ELN (DC00167).

References

Altmann, G.T.M., Dienes, Z., Goode, A., 1995. Modality independence of implicitly learned grammatical knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition 21, 899–912.

Asaad, P., 1998. Statistical learning of sequential visual patterns. Unpublished senior honors thesis. University of Rochester, New York.

Aslin, R.N., Saffran, J.R., Newport, E.L., 1998. Computation of conditional probability statistics by 8-month-old infants. Psychological Science 9, 321–324.

Aslin, R.N., Woodward, J.Z., LaMendola, N.P., Bever, T.G., 1996. Models of word segmentation in maternal speech to infants. In: Morgan, J.L., Demuth, K. (Eds.), Signal to Syntax. Erlbaum, Hillsdale, NJ, pp. 117–134.

Balaban, M.T., Anderson, L.A., Wisniewski, A.B., 1998. Lateral assymetries in infant melody perception. Developmental Psychology 34, 39–48.

Bertoncini, J., Morais, J., Bijeljac-Babic, R., McAdams, S., Peretz, I., Mehler, J., 1989. Dichotic perception and laterality in neonates. Brain and Language 37, 591–605.

Best, C.T., Hoffman, H., Glanville, B.B., 1982. Development of infant ear asymmetries for speech and music. Perception and Psychophysics 31, 75–85.

Bever, T.G., Chiarello, R., 1974. Cerebral dominance in musicians and nonmusicians. Science 185, 537–539.

Blumstein, S., Cooper, W.E., 1974. Hemispheric processing of intonation contours. Cortex 10, 146–152.

Brent, M.R., Cartwright, T.A., 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61, 93–125.

Calandra, T.S., 1998. Learning at a distance: statistical learning of non-adjacent regularities. Unpublished senior honors thesis. University of Rochester, New York.

Cartwright, T.A., Brent, M.R., 1997. Early acquisition of syntactic categories: a formal model. Cognition 63, 121–170.

Chomsky, N., 1957. Syntactic Structures. Mouton, The Hague.

Chomsky, N., 1975a. The Logical Structure of Linguistic Theory. Plenum Press, New York.

Chomsky, N., 1975b. Reflections on Language. Pantheon Books, New York.

Christiansen, M.H., Allen, J., Seidenberg, M.S., 1998. Learning to segment speech using multiple cues: a connectionist model. Language and Cognitive Processes, in press.

Christophe, A., Dupoux, E., Bertoncini, J., Mehler, J., 1994. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. Journal of the Acoustical Society of America 95, 1570–1580.

Cole, R., Jakimik, J., 1980. A model of speech perception. In: Cole, R.A. (Ed.), Perception and Production of Fluent Speech. Erlbaum, Hillsdale, NJ, pp. 133–163.

Echols, C.H., Newport, E.L., 1992. The role of stress and position in determining first words. Language Acquisition 2, 189–220.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K., 1996. Rethinking Innateness: A Connectionist Perspective on development. MIT Press, Cambridge, MA.

Fodor, J.A., 1983. The Modularity of Mind. MIT Press, Cambridge, MA.

Fodor, J.A., Bever, T.G., Garrett, M.F., 1974. The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. McGraw Hill, New York.

Gallistel, C.R., 1990. The Organization of Learning. MIT Press, Cambridge, MA.

Gelman, R., Williams, E.M., 1998. Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In: Kuhn, D., Siegler, R.S. (Eds.), Handbook of Child Psychology, Vol. 2, Cognition, Perception and Language, 5th edition (W. Damon, series editor). Wiley, New York, pp. 575–630.

Golinkoff, R.M., Alioto, A., 1995. Infant-directed speech facilitates lexical learning in adults hearing Chinese: implications for language acquisition. Journal of Child Language 22, 703–726.

Goodsitt, J.V., Morgan, J.L., Kuhl, P.K., 1993. Perceptual strategies in prelingual speech segmentation. Journal of Child Language 20, 229–252.

Harris, Z.S., 1955. From phoneme to morpheme. Language 31, 190–222.

Hasher, L., Zacks, R.T., 1984. Automatic processing of fundamental information. American Psychologist 39, 1372–1388.

Hayes, J.R., Clark, H.H., 1970. Experiments in the segmentation of an artificial speech analog. In: Hayes, J.R. (Ed.), Cognition and the Development of Language. Wiley, New York, pp. 221–234.

Hunt, R.H., Aslin, R.N., 1998. Statistical learning of visuomotor sequences: implicit acquisition of subpatterns. Poster presented at the annual meeting of the Cognitive Science Society, August, 1998, Madison, WI.

Jacobs, R.A., Jordan, M.I., Barto, A.G., 1991. Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. Cognitive Science 15, 219–250.

Jusczyk, P.W., Aslin, R.N., 1995. Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology 29, 1–23.

Jusczyk, P.W., Krumhansl, C.L., 1993. Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. Journal of Experimental Psychology: Human Perception and Performance 19, 627–640.

Jusczyk, P.W., Cutler, A., Redanz, L., 1993. Infants' sensitivity to predominant stress patterns in English. Child Development 64, 675–687.

Karmiloff-Smith, A., 1992. Beyond Modularity: A Developmental Perspective on Cognitive Science. MIT Press, Cambridge, MA.

Kelly, M.H., Martin, S., 1994. Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. Lingua 92, 105–140.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., Gerken, L.A., 1995. The headturn preference procedure for testing auditory perception. Infant Behavior and Development 18, 111–116.

Kimura, D., 1964. Left-right differences in the perception of melodies. Quarterly Journal of Experimental Psychology 16, 355–358.

Lerdahl, F., Jackendoff, R., 1983. A Generative Theory of Tonal Music. MIT Press, Cambridge, MA.

Maratsos, M., Chalkley, M.A., 1980. The internal language of children's syntax: the ontogenesis and representation of syntactic categories. In: Nelson, K. (Ed.), Children's language, Vol. 2. Gardner Press, New York, pp. 127–213.

Marcus, G., 1998. Can connectionism save constructivism? Cognition, in press.

Marler, P., 1991. The instinct to learn. In: Carey, S., Gelman, R. (Eds.), The Epigenesis of Mind: Essays on Biology and Cognition. Erlbaum, Hillsdale, NJ, pp. 37–66.

Mehler, J., Dupoux, E., Segui, J., 1990. Constraining models of lexical access: the onset of word recognition. In: Altmann, G.T.M. (Ed.), Cognitive Models of Speech Processing. MIT Press, Cambridge, MA, pp. 236–262.

Miller, G.A., Selfridge, J.A., 1950. Verbal context and the recall of meaningful material. American Journal of Psychology 63, 176–185.

Mintz, T.H., Newport, E.L., Bever, T.G., 1995. Distributional regularities of form class in speech to young children. Proceedings of NELS 25. GLSA, Amherst, MA.

Morgan, J.L., Meier, R.P., Newport, E.L., 1987. Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. Cognitive Psychology 19, 498–550.

Morgan, J.L., Newport, E.L., 1981. The role of constituent structure in the induction of an artificial language. Journal of Verbal Learning and Verbal Behavior 20, 67–85.

Morgan, J.L., Saffran, J.R., 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development 66, 911–936.

Newport, E.L., Aslin, R.N., in press. The place of statistical learning in a theory of language acquisition. In: Mehler, J., Bonatti, L., Carey, S. (Eds.), Notions of Growth and Development.

Peretz, I., 1987. Shifting ear-asymmetry in melody comparison through transposition. Cortex 23, 317–323.

Reber, A.S., 1993. Implicit Learning and Tacit Knowledge: an Essay on the Cognitive Unconscious. Oxford University Press, New York.

Saffran, J.R., 1997. Statistical learning of syntactic structure: mechanisms and constraints. Unpublished doctoral dissertation. University of Rochester, New York.

Saffran, J.R. From strings to trees: constrained statistical learning in language acquisition, submitted.

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

Saffran, J.R., Newport, E.L., Aslin, R.N., 1996b. 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., Barrueco, S., 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science 8, 101–195.

Schellenberg, E.G., Trehub, S.E., 1996. Natural musical intervals: evidence from infant listeners. Psychological Science 7, 272–277.

Seidenberg, M.S., 1997. Language acquisition and use: learning and applying probabilistic constraints. Science 275, 1599–1603.

Slobin, D.I., 1973. Cognitive prerequisites for the development of grammar. In: Ferguson, C.A., Slobin, D.I. (Eds.), Studies of Child Language Development. Holt, Rinehart and Winston, New York, pp. 407–431.

Trehub, S.E., Trainor, L.J., 1993. Listening strategies in infancy: the roots of language and musical development. In: McAdams, S., Bigand, E. (Eds.), Thinking in Sound: The Cognitive Psychology of Human Audition. Oxford University Press, London, pp. 278–327.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.