A Cross-Linguistic Study in Learning Prosodic Rhythms: Rules, Constraints, and Similarity

TODD M. BAILEY\textsuperscript{a}, KIM PLUNKETT\textsuperscript{a} and ESTER SCARPA\textsuperscript{b}

\textsuperscript{a}University of Oxford
\textsuperscript{b}University of Campinas, Brazil

INTRODUCTION

The rhythmic structure of speech reflects a hierarchical organization of the temporal sequence of speech sounds into syllables and higher level units of prosodic and syntactic structure. Knowledge of linguistic rhythm plays a central role in the ability of a competent listener to make sense of a speech stream. Since different languages employ a variety of

ABSTRACT

Differences in the learnability of linguistic patterns may be crucial in deciding among alternative learning models. This paper compares the ability of English speakers (Experiment 1) and Portuguese speakers (Experiment 2) to learn two complex rhythm patterns observed in languages with primary word stress. Subjects were familiarized with one of two rhythms during a discrimination task, followed by a recognition task which tested whether knowledge of the rhythm generalized to novel stimuli. The main findings were: (1) speakers trained on the cross-linguistically less common rhythm distinguished between novel stimuli which did or did not conform to their training rhythm, while speakers trained on the more common rhythm did not; (2) English speakers were biased more strongly than Portuguese speakers against final stress; and (3) melodies that are on a boundary between rhythm categories were treated as less prototypical than other members of the same rhythm category. The results demonstrate that knowledge of complex linguistic rhythms can be generalized after very little training, and that the less common rhythm is easier to learn even though it seems more complex. The results are compared with general-purpose exemplar-based learning models as well as abstract linguistic theories of word stress acquisition.

KEY WORDS

artificial grammar learning
prosody
rhythm
stress

\* We would like to thank Nick Chater, Ulrike Hahn, Janet Pierrehumbert, Martin Redington, and our anonymous reviewers for valuable comments on earlier drafts of this paper. Discussions with Martin Redington and Emmanuel Pothos about artificial grammar learning were also helpful. Thanks go to C. Randy Fletcher for challenging the first author to test stress theory experimentally. The first author was supported by grants from the McDonnell-Pew Center for Cognitive Neuroscience, and a grant to Kim Plunkett from the Biotechnology and Biological Sciences Research Council.

Address for correspondence: Todd M. Bailey, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, United Kingdom.
E–mail: todd.bailey@psy.ox.ac.uk
often complex rhythms, part of learning a language is to acquire knowledge of the systematic rhythmic organization of that language. An explanatory account of the knowledge implied by linguistic rhythm must distinguish between linguistically possible and impossible rhythms, and must account for differences in learnability among different patterns of linguistic rhythm. Linguistic theory has long recognized syllable stress as a reflection of underlying rhythmic structure, and the present work uses stress to investigate the learnability of complex rhythm patterns. Linguistic theories of stress have been based primarily on typologies of stress patterns observed cross-linguistically. However, the frequency of stress patterns in languages of the world does not necessarily have theoretical significance (cf. Gupta & Touretzky, 1994). Processing mechanisms can be probed more directly by assessing the learnability of different patterns. Unfortunately, there are as yet no relevant data bearing on the relative difficulty of learning different stress patterns. The present work is a first step in accumulating the relevant data.

Learning a language involves generalizing from individual linguistic tokens, to apply patterns and structure productively to novel forms. Though learning a language takes years, the ability to generalize structure from a limited set of examples can be studied in much shorter experimental tasks. Miller (1958) showed that letter strings generated by a grammar were learned more quickly and recalled more accurately than random letter strings. Reber (1969) found that after exposure to letter strings generated by a grammar, subjects performed above chance in discriminating between novel sequences that did or did not obey the rules of the grammar. Generalization of knowledge from exemplars to novel stimuli has also been shown for auditory sequences, including tone sequences and sequences of spoken syllables (Altmann, Dienes, & Goode, 1995; Manza & Reber, 1994). Also, Zwitserlood (1990) found that after a short exposure to Mandarin speech, Dutch adults could discriminate between real Mandarin words and pseudowords that violated phonotactic constraints of Mandarin.

The experiments reported below extend this general line of research in two directions. First, we test learning of rhythms like those observed in languages with primary stress. In the studies reported below, adults are exposed to one of two rhythm patterns, then tested to measure whether knowledge of the training rhythm generalizes to novel stimuli. In addition, the current research compares different attested rhythms against each other and thereby tests the relative difficulty of learning different prosodic rhythms. The main goals of this work are to determine whether it is possible for adult speakers to learn prosodic rhythms in an experimental setting, to compare the relative ease of learning different rhythm patterns, to identify the nature of the processing mechanisms underlying the learning of prosodic rhythm, and to assess the role of the speaker’s existing knowledge of prosodic patterns in learning new ones.

We proceed by providing a brief review of word stress and alternative linguistic theories of stress learning. We then describe an experimental task in which subjects learn a prosodic rhythm that is attested in the world’s languages but not in their own. We demonstrate that subjects can quickly learn prosodic rhythms and generalize them to novel stimuli. Furthermore, this ability is sensitive to prior linguistic background. We evaluate our results against proposed models of stress learning, and also against general-purpose models of generalization based on similarity to exemplars. None of the models considered is completely consistent with the experimental results.
STRESS AND LEARNABILITY

Word stress plays a special role in organizing the speech stream. Many languages exhibit alternations between stressed and stressless syllables, and a number of studies have focused on the simple distinction between syllables which are stressed and those which are unstressed (e.g., Cutler, 1986; Gerken, 1994; Pitt & Samuel, 1990). However, not all linguistic rhythms consist of simple alternations. Indeed, “the basic principle in a stress language is that only one syllable per word will receive primary stress” (Hyman, 1977, p. 38). Each word or phrase normally culminates in a single strongest syllable. English words like Mississipi and Mèditerránean have several stressed syllables, marked here with accents. One of the stressed syllables bears the primary word stress, and is perceived as more prominent than the other stressed syllables. Primary word stress potentially gives the listener an indication of the number of words in an utterance, signals their relative locations, and may also distinguish otherwise identical words (Trubetzkoy, 1969; for discussion, cf. Beckman, 1986). In general, auditory sequences with one or a few prominent elements are easier to remember than sequences with multiple prominent elements, suggesting a role for word stress in facilitating memory for words (Bell, 1977; also cf. Cutler, 1986; Frings, 1914, cited in Robinson, 1977). Stressed syllables (as opposed to unstressed ones) may be signaled acoustically with higher pitch (or more abstractly by lawful association with some pitch feature), greater intensity, longer duration, or some combination of these. The primary stressed syllable functions as a reference point for intonational contours (Selkirk, 1984), and is often acoustically marked even when other stressed syllables are not (Martin, 1972; for a detailed discussion of diagnosing the location of stress within a word, cf. Hayes, 1995, Chapter 2).

For some languages, the location of stress within a word is arbitrary, and must be memorized as part of learning each word, just as English speakers must memorize the difference in stress for the English words *Panama* (first syllable stress) and *banana* (second syllable stress). In other languages, the location of stress is determined by the phonological structure of each word, according to language-particular stress assignment rules. The acquisition of these rules is the subject of the current study.

In many languages, the location of stress within a word is sensitive to syllable structure, in a rich-get-richer fashion, so that syllables which are already perceptually prominent in a language-particular way tend to attract stress. Stress may be associated with long vowels, or syllables which end in a consonant, or syllables with a particular pitch feature, and so on, depending on the language (while in other languages syllable structure is irrelevant to stress assignment). Stress theory compartmentalizes these language-particular distinctions in the abstract notion of syllable weight, with the heavy syllables of a given language having a special attraction for stress, contrasted with light syllables which do not. In Malayalam, for example, syllable weight is determined by vowel length. Stress falls on the first syllable if it has a long vowel. If the first syllable is short, then stress falls on the second syllable if it has a long vowel. If the first two syllables are both short, stress falls on the first syllable (Hayes, 1995, p. 92). Once the distinction between light and heavy syllables is made, stress patterns can generally be described in terms of which syllables have priority for stress assignment. The pattern of Malayalam is succinctly captured by the syllable priority code “12/1L,” which is read, ‘Counting from the Left edge of the word, stress syllable 1 if heavy, otherwise stress syllable 2 if heavy, otherwise stress syllable 1’ (Bailey, 1995,
Learning prosodic rhythms

Cheremis has a more complex “23..89/2R” stress pattern (Hayes, 1981, p. 123). This syllable priority code is read, ‘Counting from the Right edge of the word, stress syllable 2 if heavy, else stress syllable 3 if heavy, and so on through the remaining syllables until the beginning of the word. If there are no heavy syllables, primary stress falls on the second syllable from the right.’

In order to evaluate stress theories, one must distinguish between logically possible rhythms, and those rhythms which either do or could occur in the stress patterns of human languages. A set of plausible word stress rhythms is listed in Table 1, along with the number of languages exhibiting each rhythm in its pattern of primary stress (in a sample of ca. 170 languages; Bailey, 1995, 1996). The present study focuses on two of these rhythms, “12..89/1R” and “23..891/2R,” from rows 7 and 17 of the table. These rhythms were chosen because they are complex enough that subjects are unlikely to discover the explicit rule used to generate a set of stimuli, and also because these two rhythms are equally complex in terms of the amount of information required to assign stress to a word. In addition, the two rhythms differ in cross-linguistic frequency. If one rhythm is more difficult to learn than the other, stress theory should account for the difference in learnability regardless of the difference in cross-linguistic frequency.

The rhythm “12..89/1R” places stress on the rightmost heavy syllable in a word, or on the rightmost syllable if there are no heavy syllables in the word. This pattern of stress is common cross-linguistically, appearing in 15 languages in the sample, including Golin and Mayan. In contrast, the rhythm “23..891/2R” is relatively rare. In words with more than one heavy syllable, the “23..891/2R” rhythm places stress on the right-most nonfinal heavy syllable. If there is a single heavy syllable in the word, that syllable receives the stress, even if it is the final syllable in the word. Otherwise, if there are no heavy syllables in the word, stress falls on the penultimate syllable. Of the languages sampled, the “23..891/2” pattern of stress is attested in only four: Sindhi, Bhojpuri, and two varieties of Hindi.

Intuitively, the “23..891/2” rhythm seems more complex than the “12..89/1” rhythm, and might therefore be expected to be less natural as a stress pattern (Hyman, 1977). However, intuition is not necessarily the relevant measure of complexity. Hyman suggests that “the less information necessary for the placement of stress, the less complex the stress placement rule” (fn. 15, p. 72). In an information-theoretic sense, the “12..89/1R” and “23..891/2R” rhythms are equally complex. The amount of information needed to assign stress to a word can be computed by counting the number of binary decisions required to determine the location of stress. To assign stress according to the “12..89/1R” rhythm, one may begin by asking “Is the first syllable heavy (counting from the right edge of the word)?” If the answer is yes, stress is assigned to the first syllable. If the answer is no, we proceed by asking, “Is the second syllable heavy?,” and so on. To assign stress according to the “23..891/2R” rhythm, one asks exactly the same questions, but the question about the first syllable is moved to the end of the sequence of questions. If heavy and light syllables are equally likely, then the average number of binary decisions required to assign either “12..89/1R” or “23..891/2R” stress to a word of n syllables is given by

\[ H = - \sum_{i=1}^{n} p_i \log_2 p_i \]

where \( p_1 = \frac{1}{2} + \left(\frac{1}{2}\right)^n \) and \( p_{i>1} = \left(\frac{1}{2}\right)^i \)
5T. M. Bailey, K. Plunkett, and E. Scarpa (cf. Quinlan 1990, p. 61, and references therein). For \( n=5 \), for example, \( H=1.42 \), the number of bits of information required to assign stress according to either the “12..89/1R” or the “... can inquire what kind of learning mechanism, linguistic or not, might account for the observed pattern of learnability.

<table>
<thead>
<tr>
<th>Index</th>
<th>Rhythm</th>
<th>Attested</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>12/1</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>12/2</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>123/1</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>12..67/1</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>12..78/1</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>12..89/1</td>
<td>15*</td>
</tr>
<tr>
<td>8</td>
<td>12..89/2</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>12..89/9</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>1298..43/1</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>198..32/1</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>13</td>
<td>21/1</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>21/2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>23/3</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>23..89/3</td>
<td>–</td>
</tr>
<tr>
<td>17</td>
<td>23..891/2</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>23..891/3</td>
<td>–</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>34..89/3</td>
<td>–</td>
</tr>
<tr>
<td>21</td>
<td>34..89/4</td>
<td>–</td>
</tr>
<tr>
<td>22</td>
<td>34..892/3</td>
<td>–</td>
</tr>
<tr>
<td>23</td>
<td>34..892/4</td>
<td>–</td>
</tr>
<tr>
<td>24</td>
<td>34..8921/3</td>
<td>–</td>
</tr>
<tr>
<td>25</td>
<td>34..8921/4</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note: The count for “12..89/1” includes two languages with the “12..89/1R” stress pattern and 13 languages with the mirror image “12..89/1L” pattern. If directionality of rhythm patterns ultimately turns out to be a significant factor in learning difficulty, left-anchored versus right-anchored patterns will have to be evaluated separately. The conflation of directionality here is confined to the discussion of attested stress patterns and does not confound the experimental results.

(cf. Quinlan 1990, p. 61, and references therein). For \( n=5 \), for example, \( H=1.42 \), the number of bits of information required to assign stress according to either the “12..89/1R” or the “23..891/2R” rhythm, averaged across all five-syllable words.

Though these two rhythms have equal complexity in an information-theoretic sense, various theoretical accounts make conflicting predictions about the relative difficulty of learning the “12..89/1R” and “23..891/2R” rhythms. By observing which rhythms are easier to learn we can further constrain theories of linguistic rhythm, requiring them not only to distinguish between linguistically possible and impossible stress patterns, but also to account for differences in learnability. More generally, we can inquire what kind of learning mechanism, linguistic or not, might account for the observed pattern of learnability.
We will consider several models of stress learning which cover a broad spectrum, from domain-specific rules to completely general models of exemplar learning. Working with representations based on metrical theory (Halle & Vergnaud, 1987; Hayes, 1981; Liberman, 1975; Prince, 1983), Dresher and Kaye (1990) and Nyberg (1990) assume a small, fixed set of domain-specific rules for building rhythm-defining constituent structures, with a few binary parameters accounting for different stress patterns. Within optimality theory (Prince & Smolensky, 1993), Tesar (1997a, b) assumes similar constituent structures, but employs a fixed set of domain-specific constraints instead of rules, with differences in constraint ranking accounting for different stress patterns. Bailey (1995) assumes domain-specific constraints on competition between syllables for prominence, with different (continuous-valued) constraint weighting accounting for different stress patterns. Gupta and Touretzky (1994) assume a connectionist stress-assignment network with minimal domain-specific structure. Because learning stress patterns might be considered a special case of category learning, we will also consider general-purpose learning models based on whole-item or fragment similarity between test and training items (Meulemans and Van der Linden, 1997; Redington & Chater, 1996; Redington, 1996; Servan-Schreiber & Anderson, 1990).

**Rule-Based Metrical Theory**

According to metrical theory (Halle & Vergnaud, 1987; Hayes, 1981; Liberman, 1975; Prince, 1983), stress patterns reflect an abstract rhythm structure which organizes syllables into hierarchical constituents. The basic unit of rhythm above the syllable is the metrical foot, which represents a prominence relation between the syllables within it. The head of each foot is relatively prominent or stressed, relative to nonhead syllables. Feet are grouped into a metrical word, and the head of one foot serves as the head of the whole word and bears primary word stress. Languages vary in the way they group syllables into feet, and also in how the head foot within a word is determined.

Dresher and Kaye (1990) and Nyberg (1990) propose models of stress acquisition which adopt a bottom-up view of stress assignment. Universal stress rules operate on a syllabified word. Syllables are grouped into feet, and each foot has a designated head syllable. Feet are then grouped into a phonological word, with one foot serving as the head. Foot heads are stressed, and the head syllable of the head foot bears primary word stress. Eleven binary parameters are posited to account for variation in stress patterns between languages. According to Dresher and Kaye, these eleven parameters can be combined to yield 216 distinct stress systems. The task of a language learner is to discover the parameter settings for the target language.

The parameter which primarily concerns us here is extrametricality, which determines whether the final syllable of a word is skipped over by the stress assignment process. It is the setting of this parameter alone which distinguishes the “12..89/1R” and the “23..891/2R” stress patterns. For the “12..89/1R” rhythm, extrametricality must be [off], while the opposite parameter setting, [on], is required for the “23..891/2R” rhythm. Figure 1 shows how the setting of the extrametricality parameter affects the location of stress when syllables are grouped into a single, unbounded foot, whose head is on the right. For simplicity, the effects of syllable weight are not shown in Figure 1. Parentheses show foot boundaries, asterisks identify syllables and foot heads. Extrametricality renders the final syllable invisible to stress assignment rules.
T. M. Bailey, K. Plunkett, and E. Scarpa

Although Nyberg (1990) adopts the same set of parameters as Dresher and Kaye (1990), he adopts different assumptions about the default values of those parameters. Since it takes time and experience with a stress pattern to set parameters, it takes longer to learn stress patterns which require more parameters to be changed from their default settings. Under the treatment of extrametricality taken by Dresher and Kaye (1990), extrametricality is [on] by default, so the “23..891/2R” rhythm should be easier to learn than the “12..89/1R” rhythm. Interestingly, this is the only model which seems to explicitly predict the pattern of results observed in our studies. In the Nyberg model, extrametricality is [off] by default, so the “23..891/2R” rhythm is more difficult for this model to learn than the “12..89/1R” rhythm.

Stress in Optimality Theory

Working within optimality theory, Tesar (1997a, b) proposes an algorithm for learning stress by reranking constraints on stress. Instead of applying rules to build metrical structure, optimality theory (Prince & Smolensky, 1993) evaluates the set of possible structures against a set of universal constraints which are ranked in importance in a language-specific hierarchy. Whichever metrical structure violates the fewest high-ranking constraints is selected as the winner. Of the 12 universal constraints Tesar posits to account for stress assignment, the five in Table 2 are critical to the analysis of the “12..89/1R” and “23..891/2R” patterns.

Optimality theory does not require that constraints be satisfied, just that of all surface forms that might possibly be generated for a given word, the winning form violates fewer high-ranked constraints compared to any of the competing forms. For example, Parse-Syllable and NonFinal will always conflict, and the conflict will be resolved in favor of whichever constraint is ranked higher in a particular language. In effect, NonFinal enforces final syllable extrametricality, but only if it is highly ranked relative to other constraints on metrical structure. Using the conventional notation X >> Y to mean X is ranked higher

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight-to-Stress</td>
<td>heavy syllables should be stressed</td>
</tr>
<tr>
<td>Parse-Syllable</td>
<td>each syllable should belong to some foot</td>
</tr>
<tr>
<td>NonFinal</td>
<td>the final syllable should not belong to any foot</td>
</tr>
<tr>
<td>All-Feet-Right</td>
<td>feet should be as close as possible to the right edge of the word</td>
</tr>
<tr>
<td>All-Feet-Left</td>
<td>feet should be as close as possible to the left edge of the word</td>
</tr>
</tbody>
</table>

Figure 1

Effect of extrametricality on foot construction in a five syllable word.
than Y, the “12..89/1R” stress pattern requires the constraints to be (partially) ranked so that Weight-to-Stress >> All-Feet-Right >> {NonFinal, All-Feet-Left, Parse-Syllable} (curly brackets group constraints whose ranking relative to each other does not matter). The “23..891/2R” stress pattern requires the ranking Weight-to-Stress >> NonFinal >> All-Feet-Right >> {All-Feet-Left, Parse-Syllable}.

The number of different (partial) rankings of twelve constraints is enormous. The number of distinct stress systems produced by this set of constraints is probably much smaller, perhaps on the order of hundreds, but a precise count is difficult to determine. The task of a language learner is to arrive at the right constraint ranking, and Tesar proposes an algorithm of constraint demotion to accomplish this. Observed words are compared with predictions generated by the learner’s current grammar (i.e., constraint hierarchy). If the predicted stress deviates from the observed stress, an appropriate set of constraints is moved lower down within the hierarchy, bringing the learner’s grammar closer to the target grammar. Multiple constraints can be demoted simultaneously, but only one new level can be added to the hierarchy on each round. Languages which require more rounds of constraint demotion will take longer to learn.

The number of rounds of constraint demotion required to reach the target grammar will depend on the initial constraint ranking. If all constraints are initially unranked, at least two rounds of demotion are required to arrive at the three-level ranking hierarchy required for the “12..89/1R” pattern, and at least three rounds of demotion are required to arrive at the “23..891/2R” pattern. Under these assumptions, the “12..89/1R” stress pattern will be easier to learn than the “23..891/2R” pattern, because the “23..891/2R” pattern requires more rounds of constraint demotion. Tesar points out, however, that a number of attested stress patterns are unlearnable under these conditions. In order to expand the range of attested stress patterns which may be learned by the constraint demotion algorithm, Tesar proposes a more structured initial constraint hierarchy in which Weight-to-Stress >> Parse-Syllable >> {NonFinal, All-Feet-Right, All-Feet-Left}. Starting from this hierarchy enables the demotion algorithm to arrive at the correct grammar for a number of stress patterns which are otherwise unlearnable. At least one round of demotion is required for “12..89/1R” stress, and at least two rounds of demotion are required for “23..891/2R” stress. Again, the “12..89/1R” pattern should be easier to learn than the “23..891/2R” pattern on this account.

**Nonmetrical Constraints Model of word stress**

Both the rule-based account and the optimality theory account of word stress emphasize the role of metrical feet in determining the location of word stress. In contrast, the nonmetrical constraints model (Bailey, 1995) views word stress assignment as a competition between syllables for prominence, emphasizing direct relations between syllables and word stress. Competition for prominence does not do away with metrical feet, but reduces their role in determining the location of word stress, especially in unbounded stress patterns like “12..89/1R” and “23..891/2R.”

The nonmetrical constraints model is based on the observation that stress patterns vary along two dimensions involving directionality and edge effects. Stress may be assigned from either the left or the right edge of a word, as seen in the contrast between the “1L” pattern of initial stress in Latvian and the “1R” pattern of final stress in Weri. Moreover,
the “12..89/1L” pattern of Tibetan (placing stress on the first heavy syllable) and the “12..89/1R” pattern of Aguacatec Mayan (placing stress on the last heavy syllable) can be seen as occupying intermediate points on a left-right directionality continuum between the extremes of “1L” and “1R.” Many stress patterns also involve local edge effects whereby syllables close to the edge of a word (or other morphosyntactic constituent) exhibit a special attraction for or avoidance of stress. These edge effects are seen in the progressive retraction of stress observed in the “1R,” “2R,” and “3R” (final, penultimate, and antepenultimate stress) of Weri, Polish, and Paamese. The nonmetrical constraints model incorporates directionality and edge biases directly as weighted constraints on a competitive process which determines the location of stress. Various combinations of values for these parameters produce all the stress patterns mentioned above, including those listed in Table 1. Figure 2 shows the parameter regions for the “12..89/1R” and “23..891/2R” stress patterns of interest. The horizontal axis represents directionality bias (positive values favor each syllable over others on its left). The vertical axis represents edge bias in favor (positive values) or against (negative values) stress near the end of a word.

For the nonmetrical constraints model, learning a stress pattern involves setting real-valued parameters which determine the relative importance of directionality and edge bias constraints on the competition for prominence. If parameters are changed incrementally, then the learnability of stress patterns will be affected by the difference between initial parameter values and the values required for the target stress pattern. Also, if performance factors such as noise limit the precision of mental representations (e.g., parameter values or competition calculations), then learnability may also be affected by the size of the parameter region producing each distinct stress pattern. The “12..89/1R” pattern involves a smallish directionality bias towards the right, in combination with little or no edge bias. 

Figure 2
Parameter regions in the nonmetrical constraints model of word stress, for stress patterns including “12..89/1R” (center) and “23..891/2R” (bubble near lower right).
The “23..891/2R” pattern has a similar directionality bias towards the right, but also involves a moderate local edge bias against stress at the end of words. If a learner starts with no directionality or edge biases (i.e., these parameters are initially zero), then the “23..891/2R” pattern should be harder to learn because it is farther from the origin in parameter space. Moreover, the size of the parameter region for “23..891/2R” is much smaller than the size of the region for “12..89/1R” (cf. Bailey, 1995, Chapter 4). If fine-grained parameter settings are more difficult than coarse-grained settings, again, the “23..891/2R” pattern should be more difficult for learners to master.

Perceptron Model of Stress

Gupta and Touretzky (1994; G&T) propose a perceptron model of stress acquisition, in which a two-layer connectionist network learns to associate syllable weight strings with stress patterns. This model dispenses not only with metrical constituent structures, but endeavors to account for stress patterns with a simple general-purpose learning device with no domain-specific structure built in. The model consists of an input array connected to a single output. A word is processed one syllable at a time by sliding it through the input array, with each position of the input array representing the weight of a syllable. At each time step, the desired output is the stress level of the middle syllable of the input array. Learning proceeds by making small adjustments to the weights of the connections between input and output, reducing the discrepancy between the actual and desired outputs for the current syllable. Gupta and Touretzky report the relative difficulty the perceptron model has in learning various stress systems, and relate differences in learnability to aspects of metrical stress theory. The “12..89/1R” stress pattern was one of the hardest stress patterns for the perceptron to learn.

Since G&T did not include the “23..891/2R” stress pattern in their simulations, we implemented the perceptron model using TLEARN (Plunkett & Elman, 1997) in order to directly compare the learnability of the “12..89/1R” and “23..891/2R” rhythms with respect to this model. Our simulations replicated G&T’s finding that the “12..89/1R” pattern is very difficult but learnable. However, the perceptron model failed to learn the “23..891/2R” pattern to criterion, even after twice as many learning trials. In three repetitions from different initial conditions and random training orders, the “12..89/1R” rhythm was consistently learned in less than 320,000 training trials; the “23..891/2R” rhythm was not learned even after 640,000 trials. The apparent unlearnability of the “23..891/2R” rhythm is a problem for the perceptron model, since several languages are attested as having this stress pattern. In any event, the perceptron model clearly predicts the “23..891/2R” rhythm should be more difficult to learn than the “12..89/1R” rhythm.

In summary, the Dresher and Kaye model predicts that the “23..891/2R” rhythm should be easier to learn than the “12..89/1R” rhythm. The other learning models predict the opposite order of difficulty. An explicit measure of the relative learnability of these rhythms will put these predictions to the test.

Length, Pitch, and Stress

In general, the acoustic correlates of stress include pitch, intensity, and duration. The realization of stressed syllables varies from language to language, and even within a
language the realization of stress depends on the phonological and syntactic context. This variability in the realization of stress poses a challenge for anyone constructing stimuli for studying different patterns of stress. However, linguistic theory has long recognized stress as a reflection of abstract rhythmic timing relationships (cf. Hayes, 1995; Lenneberg, 1967; Liberman, 1975; Martin, 1972). As Hayes noted, “rhythm in general is not tied to any particular physical realization” (p. 8). Bell (1977) demonstrated the perception of rhythm over a wide range of noises, including tones and hisses. This suggests that one can reasonably abstract away from the details of phonological syllables when investigating stress patterns, and use stimuli composed of simple tones, concentrating on the timing and prominence relationships at the heart of linguistic rhythm. Though languages tolerate variability in the actual realization of stress, alterations in pitch are perceptually more salient than changes in intensity or duration (Fry, 1955, 1958, and other works cited in Hayes, 1995, p. 5). Accordingly, the stimuli employed in the experiments below are simple melodies of short and long tones, with one tone in each melody having a higher pitch than the others.

It should be noted that any rhythm contrived by manipulating length and pitch is ambiguous between several possible linguistic interpretations. This is because syllable length and pitch play several roles in phonological structure. In many languages stressed syllables tend to be longer than unstressed ones, but some languages also use length phonemically to contrast short and long syllables. In the case of pitch, in addition to being a cue for stressed syllables, and in tone languages being used phonemically, pitch is also used to signal intonational contours across phrases. Length might well be interpreted as a cue for phonemic syllable length, and pitch as a cue for stress. In that case, the stimuli employed in the present study would be perceived as pseudo words composed of phonemically short or long syllables, one of which is stressed. Alternatively, length might be interpreted as a cue for stress, and pitch as a reflection of phrasal intonation. In this case, our stimuli would be perceived as pseudo phrases of stressed and unstressed syllables, with a single intonational peak. In the latter case, specific theories of stress assignment, per se, may not be directly relevant. However, both interpretations involve suprasegmental organization of syllables into higher level linguistic structure. Our main interest is to investigate the learnability of such prosodic patterns and to compare the adequacy of general-purpose learning theories with domain-specific models incorporating elements of linguistic theory.

**EXPERIMENT 1**

Experiment 1 was designed to evaluate whether subjects would generalize learning from training melodies to novel melodies, and if so, to test whether the “12..89/1R” rhythm or the “23..891/2R” rhythm generalized more readily. Stimuli consisted of simple melodies of short and long tones, with one tone at a higher pitch than the other tones in the same melody. A discrimination task was used for familiarization: subjects heard pairs of melodies and judged whether the two melodies were the same or different. All the training melodies

---

1 We are grateful to Janet Pierrehumbert for bringing this ambiguity to our attention, and for suggesting that the ambiguity might be eliminated by elaborating stimuli with a beat marking the intended rhythm (one beat per short tone, two beats per long tone).
for a given subject exemplified the same target rhythm. In a subsequent recognition task, subjects were asked to categorize melodies according to whether or not they had heard them during the training phase. Generalization was assessed by comparing false recognition rates for novel melodies which had the same rhythm as the training melodies versus those which did not have the same rhythm. Baseline performance is often better than chance on such tests, so it is important to have a control group against which to compare performance. The present experiments assess relative performance, with the two rhythm conditions acting as controls for each other.

Method

Subjects. Twenty-four English-speaking subjects from the University of Oxford participated in the study. Participants included an arbitrary combination of males and females from a pool consisting of about two-thirds females. Most subjects were between 18 and 30 years of age.

Stimuli. Sequences of five pure tones were used as training and testing melodies. Tones varied in length (short or long) and pitch (low or high). Within each melody, one tone was assigned stress by giving it a high pitch (200 Hz). All other tones were given low pitch (150 Hz). Short tones were 125 ms in duration, including a 20 ms onset in which the amplitude of the tone increased linearly from 0 to the maximum amplitude, and a complementary 20 ms offset in which the amplitude decreased linearly to 0. Long tones were twice as long as short tones, that is 250 ms in duration, including the same 20 ms onset and offset. These tone lengths (125 and 250 ms), are comparable to the length of short and long syllables at slower natural speaking rates. The stimuli were generated by computer and stored as digitized sound waves at a sampling rate of 22 kHz.

In discussing melodies, short and long tones will be represented with the symbols o and X, respectively. High tones will be underlined. Thus, XXooo denotes a melody composed of two long tones followed by three short tones, the last of which carries a high pitch. Ignoring pitch, short and long tones can be combined to form 32 different melodies of length five (ooooo, oooXo, oooXX, etc.). Examples of the “12..89/1R” rhythm include ooooo, oooXo, oooXX, and for the “23..891/2R” rhythm, oooo, oooX, oooXX. Target melodies were generated by assigning the high tone within a melody according to the target rhythm. Among the target melodies for each rhythm, we identified the minimal set of key melodies which would uniquely distinguish each target rhythm from all other patterns in the typology of stress patterns produced by the nonmetrical constraints model. For the “12..89/1R” rhythm, the key melodies are ooooo, Xoooo, XoooX, and XXXooo. The first form requires that a final short syllable will win the competition for stress against all other short syllables. The second and fourth key melodies require that a final short syllable lose the competition to a long syllable elsewhere in the word. The third and fourth key melodies establish a rightward bias among long syllables. Any combination of directionality and edge bias parameters which correctly assigns stress to these four key melodies will necessarily assign “12..89/1R” stress to all combinations of short and long syllables up to five-syllables in length, at least. For the “23..891/2R” rhythm, the key melodies are oooo, oooX, oooXX, XoooX, XooXo, and XXooo.

Training set. A randomized set of 16 different training melodies was constructed for each subject. All 16 training melodies exemplified the same target rhythm. In order to guarantee
that each subject had sufficient information about the training rhythm, the training set for each subject included all the key melodies for that subject’s training rhythm. These key melodies were augmented by other target melodies chosen at random to yield a set of 16 different training melodies for each subject. The 16 training melodies were organized into a block of 16 melody pairs. Half of the melody pairs were matched pairs, consisting of a single melody repeated twice. The other melody pairs were unmatched, consisting of two different melodies (both of which exemplified the target rhythm). Assignment of melodies to matched or unmatched pairs was random, but each melody occurred once as the first member of a pair and once as the second member of a pair. The order of matched and unmatched pairs within the block was also random, except care was taken to ensure that each melody (save one) appeared as the first member of a pair before appearing as the second member of a pair. Making each melody appear as the first member of a pair before appearing as the second member of a pair precludes a subject from recognizing the first member of a pair and inferring that the second member of the pair must be different, without actually listening to the second member of the pair. The complete training set for each subject consisted of three such blocks, each containing 16 pairs of melodies. The same 16 melodies were used in each block, with different randomized melody pairs within the blocks. In total, each subject listened to 48 pairs of melodies during the training phase, and heard each of 16 melodies six times.

**The test set.** The test set for each subject included 24 melodies at three levels of familiarity: old, new, and different. The old melodies were a subset of each subject’s training melodies, so the old melodies were exemplars of the target rhythm and were heard by the subject during training. The old melodies were subdivided into key melodies and nonkey melodies, as described above. For each subject, the fixed set of key melodies was augmented by additional melodies chosen at random, to yield a (potentially) different set of eight old melodies for each subject.

Eight new melodies for each subject were chosen at random from the 16 target melodies which were not included in the training set for that subject. The new melodies therefore exemplified the target rhythm, but were different from the exemplars included in the set of training melodies.

For each of the new melodies, a maximally similar different melody was created so as to make corresponding new and different melodies as similar as possible. Each different melody had the same combination of short and long tones as the corresponding new melody — only the location of the high pitch within the melody was different. The choice of alternative stress location for different melodies was based on the competition process of the nonmetrical constraints model of stress (Bailey, 1995). The competition between syllables in this model produces a rank ordering of stress locations in each word. Ordinarily, stress is assigned to the syllable at the highest-ranked location, and this was how the high pitch was assigned in each old and new melody. To create corresponding different melodies, high pitch was assigned to the second-higher-ranked location — the next best candidate location for stress. A subject assigned to the “12..89/1R” rhythm, for example, might be tested on the new melody oooXo, with a high pitch on the penultimate tone. This same subject would also be tested on the corresponding different melody oooXg, with a high final tone. Both melodies have the same combination of short and long tones, but the latter melody does not conform to the “12..89/1R” rhythm.
Design. The 24 subjects were divided into two groups of 12. One group was trained on exemplars of the “12..89/1R” rhythm, and was tested on corresponding target and nontarget test melodies. The other group of subjects was trained on exemplars of the “23..891/2R” rhythm, and was tested on corresponding test melodies. Response keys were counterbalanced, with half the subjects in each group pressing 1 and 3 to indicate yes and no, respectively. The remaining subjects had the reverse correspondence, pressing 1 and 3 to indicate no and yes, respectively.

Procedure. Subjects sat in front of a computer in a quiet room. The computer presented instructions and response cues on the screen, and played sound stimuli through headphones. At the beginning of the experiment, subjects read a screen with a version of the following instructions:

During the first part of the experiment you will hear short tone melodies, two at a time. Listen to each pair of melodies and decide whether the melodies are the same or different. If the melodies are the same, press 1 (3) as quickly as possible. If the melodies are different, press 3 (1) as quickly as possible. The computer will remind you to press 1 or 3. After you press 1 or 3 you will hear two more tone melodies, and so on. Press any key to begin.

Subjects then heard the pairs of melodies in the training set, and responded by pressing 1 or 3 on the numeric keypad. The two melodies in each pair were separated by 500 ms of silence. As the second melody in each pair began, the computer screen displayed a version of the following reminder:

Same?
1=Yes 3=No

The reminder was erased as soon as the subject pressed 1 or 3, and the next pair of melodies began 1 s later. During the training phase, subjects did not know they would be given a memory test. Upon completion of the training phase, a version of the following instructions appeared on the screen:

The first part of the experiment is complete! The next part of the experiment is a memory test. You will hear one melody at a time. Listen to each melody and decide whether you heard the melody during the first part of the experiment. If you heard the melody during the first part of the experiment, press 1 (3) as quickly as possible. If you did not hear the melody during the first part of the experiment, press 3 (1) as quickly as possible. The computer will remind you to press 1 or 3. Press any key to continue.

Subjects then heard test melodies one at a time. Subjects used the same keys that they used during the training phase to indicate yes and no. Immediately after each melody, the computer displayed a version of the following reminder:

Heard before?
1=Yes 3=No

The reminder was erased as soon as the subject pressed 1 or 3, and the next melody began 500 ms later. Reaction times were measured from the onset of the reminder prompt following the end of the test melody.

Results

Results of the training and test phases will be discussed in that order. An alpha level of .05 was used for all statistical tests.
Training analysis. Responses for the training phase were analyzed to verify subjects’ attention to the task, and to test for differences between the two training groups. As intended, subjects found the discrimination task of the training phase quite easy. Overall, the rate of correct classifications of melody pairs as same or different was 93%. The accuracy (proportion correct) and reaction times (median RTs) for each subject were analyzed to test for differences between matched and unmatched trials, and differences between the two training groups. In the analysis of accuracy, subjects were more accurate on matched than on unmatched trials, 97% versus 90%, respectively, $F(1,22)=11$, $p<.005$, 2×2 ANOVA. There was no main effect of training rhythm, nor an interaction between training rhythm and matched versus unmatched trials, $F<1$ for both. In the analysis of reaction times, subjects in the “12..89/1R” group responded more slowly than subjects in the “23..891/2R” group, $F(1, 22)=8.76$, $p=.007$. For the “12..89/1R” group, the mean RTs were 1001 ms for matched trials and 1065 ms for unmatched trials, compared with 889 ms and 862 ms for the “23..891/2R” group (SE: 21, 42; 48, 59). There was no significant difference for matched and unmatched trials, $F<1$, nor an interaction between training rhythm and matched versus unmatched trials, $F(1,22)=1.77$, $p=.20$.

The high accuracy for both rhythm groups suggests that subjects attended equally well to the training melodies. The difference in reaction times between the two rhythms might suggest a difference in processing speed for the two rhythms. Alternatively, this difference could simply reflect a difference in subject reaction times (especially since the difference in reaction times was not replicated in Experiment 2). In future, it might be helpful to include a reaction time test which was neutral with respect to rhythm type. A third possibility which was not supported by post hoc analysis is that the difference in reaction times was an accidental artifact of the random set of melodies chosen for each subject, or a reflection of the fact that the “23..891/2R” rhythm often places stress earlier in the melody compared to the “12..89/1R” rhythm. In order to test whether the training items favored earlier discrimination for subjects in the “23..891/2R” condition compared to subjects in the “12..89/1R” condition, a discriminability point was assigned to each pair of training melodies by identifying the first tone (1–5) at which two melodies differ, or 6 if the melodies were identical. The mean discriminability points were 3.78 and 3.81 for the “12..89/1R” and “23..891/2R” training groups, respectively. The two sets of discriminability points were not significantly different in a Wilcoxon rank sum test, $W(576,576)=330328$, $p=.74$.

Test analysis. The test phase was sufficiently difficult that subjects thought they were simply guessing. Results are summarized in Figure 3. The main analysis tested whether subjects remembered individual training melodies, whether the rhythm pattern of the training melodies influenced responses to novel melodies, and whether the two training groups differed. In sum, old melodies enjoyed a recognition advantage over new melodies, especially for subjects trained on the “23..891/2R” rhythm. Also, only subjects trained on the “23..891/2R” rhythm generalized their training rhythm to novel stimuli, distinguishing between new and different melodies. Results for the proportion of yes responses will be presented first, followed by reaction time results. Finally, an analysis of prototype differences among old melodies will be presented. A summary of the most specific tests appears in the top half of Table 3.

Yes/No analysis. Mean rates of yes responses are shown in Figure 3a. Although only a third of the test stimuli were old melodies from the training phase, subjects responded yes more
often than no, classifying more melodies as familiar than unfamiliar. Inspection of the results suggests that old melodies were more likely than new or different melodies to be classified as familiar. For subjects in the “23..891/2R” rhythm group, different melodies were much less likely than new melodies to be classified as familiar. However, for subjects in the “12..891/1R” group, different melodies were just as likely as new melodies to be classified as familiar. These observations are supported by statistical tests, described below. An additional mini-experiment tested whether the difference in performance between the two rhythm groups might be due to a difference in perceptual discriminability among the test items chosen for each group. This hypothesis was not supported.

No serious departures from normality or deviations in variance were observed in the data, and analysis proceeded by performing a 2×3 ANOVA on the rates of yes responses,
with training rhythm (between subjects) and melody familiarity (within subjects) as factors. There was a significant main effect of familiarity and an interaction between familiarity and training rhythm, $F(2,44)=32, p<.001$ for the main effect, $F(2,44)=25, p<.001$ for the interaction. The main effect of training rhythm missed significance, $F(1,22)=3.7, p=.07$.

In order to test whether the familiarity effects reflected memory for individual training melodies or generalization to novel melodies, new melodies were used as a reference category to obtain a memorization contrast (old vs. new) and a generalization contrast (different vs. new). Effects of memorization and generalization were tested in 2×2 ANOVAs, with training rhythm as the other factor. In the memorization ANOVA, old melodies were significantly more likely than new melodies to be classified as familiar, $F(1,22)=5, p=.03$. There was no main effect of training rhythm, nor an interaction between training rhythm and the memorization effect, $F<1$ for both. In the generalization ANOVA, main effects and the interaction were all significant, $F(1,22)=24, p<.001$ for generalization (different vs. new), $F(1,22)=7.6, p=.01$ for training rhythm, $F(1,22)=32, p<.001$ for the interaction.

Finally, to verify that the generalization effects were due to responses of the “23..891/2R” group, separate $t$-tests for each training group compared responses to new and different melodies. Subjects in the “23..891/2R” group accepted new melodies as familiar significantly more often than different melodies, $t(11)=9, p<.001$. Subjects in the “12..891/2R” group did not generalize their training rhythm, but accepted different melodies as familiar just as often as new melodies, $t(11)=0.4, p=.67$.

The pattern of results shown in Figure 3a, together with the tests described above, suggest that subjects trained on the “23..891/2R” rhythm responded in accord with their training rhythm while subjects trained on the “12..891/2R” rhythm did not. Subjects trained

### TABLE 3

**Summary of most specific comparisons in Experiments 1 (English) and 2 (Portuguese).** For each group of subjects, shows whether fraction of yes responses (% Yes) and reaction times for yes responses (RT) supported effects of generalization (different vs. new), memorization (old vs. new), and prototype structure (old key vs. nonkey). *= significant effect, √= nonsignificant trend, — = data inconsistent with effect.

<table>
<thead>
<tr>
<th>Language</th>
<th>Rhythm</th>
<th>Measure</th>
<th>Generalization</th>
<th>Memorization</th>
<th>Prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>12..89/1R</td>
<td>% Yes</td>
<td>—</td>
<td>√</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>—</td>
<td>√</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>23..891/2R</td>
<td>% Yes</td>
<td>√</td>
<td>√</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>√</td>
<td>√</td>
<td>—</td>
</tr>
<tr>
<td>Portuguese</td>
<td>12..89/1R</td>
<td>% Yes</td>
<td>√</td>
<td>—</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>—</td>
<td>—</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>23..891/2R</td>
<td>% Yes</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Note: *Main effect of memorization is significant when both training groups are considered.
on the “23..891/2R” rhythm were much better than their counterparts at rejecting different melodies as unfamiliar. Since different items were designed by rule to be maximally similar to the new items for a particular rhythm, it is conceivable that the apparent generalization effect reflects a difference in perceptual similarity of the test items for the two rhythms. If subjects could not hear the difference between “12..89/1R” new and different items, it would not be surprising that they treat them as the same in the familiarity judgment task. In order to eliminate this potential explanation it is necessary to assess the perceptual discriminability of new and different test items for the two groups.

Impressionistically, it is not difficult to hear the difference between new and different items for either rhythm pattern. To test discriminability more formally we asked six additional subjects who had not been exposed to the training stimuli to perform a discrimination task much like the training task of the original study. Each subject heard all pairs of corresponding new and different melodies from both rhythms, and heard each new and different item paired with itself. Performance was slightly better for the “12..89/1R” discriminations than for the “23..891/2R” discriminations. Of melody pairs relevant for the “12..89/1R” and “23..891/2R” rhythms, 71% and 64%, respectively, were correctly classified as same or different by all six subjects. The difference in discriminability was not significant, $W(72,76)=5519$, $p=.47$, Wilcoxon test of rank sums. The trend in favor of the “12..89/1R” rhythm here contrasts with the experimental results which showed subjects trained on the “23..891/2R” correctly rejecting different melodies while subjects trained on the “12..89/1R” rhythm did not. In short, no evidence was found to support the hypothesis that the experimental results reflect a difference in perceptual discriminability within the test melodies for the two rhythm conditions.

**Reaction time analysis.** Reaction times for yes and no responses were analyzed separately. Average reaction times for yes responses are shown in Figure 3b (reaction times for no responses revealed no clear differences for either training group). In order to minimize potential effects of skewed distributions or outlying RTs, analyses were performed on the median scores for each subject at each level of familiarity. Where average RTs are reported, they are averages across these median RTs. The use of median scores is especially appropriate since the number of RTs for each subject in each condition is small—indeed, subjects who gave all yes (no) responses for the eight melodies at one level of familiarity have no corresponding RTs for analysis of no (yes) responses. Inspection of the reaction times for subjects in the “23..891/2R” group shows that yes responses were fastest for old melodies, slowest for different melodies, and intermediate for new melodies. The reaction times for subjects in the “12..89/1R” group show no clear trend. These observations are supported by the statistical tests described below.

Because of large differences in variance between cells, nonparametric tests were used to compare reaction times. Subjects were included in each analysis only if they had at least one score for each level of melody familiarity involved in the test. In order to test whether familiarity had any effect, the reaction times for each training group were subjected to Friedman’s test of ranks for related samples. For subjects trained on the “23..891/2R” rhythm, a test of the null hypothesis that all three levels of familiarity were the same was significant for yes responses, $\chi^2(2)=6.2$, $p=.05$. The test was not significant for no responses, $\chi^2(2)=1.3$, $p=.53$. For subjects trained on the “12..89/1R” rhythm, there was
no significant effect of familiarity on RTs, $\chi^2(2) = 3.8$, $p = .15$ for yes responses, $\chi^2(2) = 4.2$, $p = .12$ for no responses.

In order to test whether the familiarity effect observed for yes responses of subjects in the “23..891/2R” group was due to memory for training melodies or generalization of the training rhythm, new melodies were used as a reference category for comparison with old and different melodies. Neither memorization nor generalization on its own was significant, $Z=1.4$, $p = .17$ for memorization, $Z=1.3$, $p = .20$ for generalization, Wilcoxon matched pairs signed ranks. In order to compare the training groups to each other, two new variables were created by subtracting each subject’s median new RT from their median old and different RTs. These variables measure the RT advantage or disadvantage of old and different melodies compared to new melodies, that is memorization and generalization, respectively. Only yes responses were included in this analysis. There was no significant difference between the two rhythm groups in the memorization effect, $W=95$, $Z=1.1$, $p = .29$, Wilcoxon test of rank sums. The generalization effect was significantly greater for subjects trained on the “23..891/2R” rhythm compared to subjects trained on the “12..89/1R” rhythm, $W=138$, $Z=1.97$, $p = .05$.

Prototype effects. Performance on almost any classification task is faster and more accurate for more prototypical members of a category compared to less prototypical members (Posner & Keele, 1968; Rosch, 1978). In the present case, key melodies were identified as critical for distinguishing one stress from another, that is, given the parameters of the nonmetrical constraints model of stress assignment, the key melodies define the boundaries between rhythm categories. If the relevant psychological space matches the structure posited by the nonmetrical constraints model, key melodies should lie near the fringes of their rhythm categories and should be relatively difficult to identify as members of the category. Nonkey melodies within the same rhythm category should be easier to classify. To test this hypothesis, old key melodies were compared to old nonkey melodies. Analysis of the proportion of yes responses for each training group will be presented first, followed by analysis of reaction times. These analyses show a recognition advantage for nonkey melodies compared to key melodies, though the advantage is statistically significant only for the “23..891/2R” group.

Nonkey melodies were more likely than key melodies to be classified as familiar, and this pattern was observed in the responses of most subjects. For subjects in the “23..891/2R” group, 88% of nonkey melodies were classified as familiar, compared to 69% of key melodies. Of the 12 subjects, the number showing a greater bias for nonkey melodies, key melodies, or no difference, was nine, two, and one, respectively, and this difference was significant in a Wilcoxon matched-pairs test, $Z=2$, $p = .02$, one-sided. For subjects in the “12..89/1R” group, 75% of nonkey melodies were classified as familiar, compared to 69% of key melodies. Of the 12 subjects, the number showing a greater bias for nonkey melodies, key melodies, or no difference, was five, four, and three, respectively. The difference between key and nonkey melodies for these subjects is consistent with a nonkey bias, but is not significant, $Z=0.8$, $p = .2$. To test for differences between the two training groups, a single prototype bias score was computed for each subject by subtracting the proportions of yes responses for key and nonkey melodies. The prototype bias scores of the two training groups were not significantly different from each other, $Z=1$, $p = .3$, two-sided.
To test for prototypicality effects in RTs, median scores were determined for each subject’s responses to old key and nonkey melodies. There were not enough subjects with no responses to both key and nonkey melodies, so only yes responses were subjected to this analysis. For subjects trained on “23..891/2R,” yes responses for key melodies were significantly slower (851 ms, compared to nonkey melodies (603 ms), Z=1.88, p=.03, 1-tailed, Wilcoxon matched pairs signed ranks test. For subjects trained on “12..89/1R,” the difference between key and nonkey melodies (1084 and 909 ms respectively) was not significant, Z=1.07, p=.15.

Discussion

Differences were found between the rhythms “12..89/1R” and “23..891/2R,” during familiarization as well as during the test phase. In the familiarization phase, subjects in the two rhythm groups were equally accurate in discriminating pairs of melodies. Correct classifications were generally faster than incorrect classifications. However, responses tended to be faster for subjects listening to melodies with the “23..891/2R” rhythm than for subjects training on the “12..89/1R” rhythm. This difference in reaction times might suggest that melodies with the “23..891/2R” rhythm were processed more quickly than melodies with the “12..89/1R” rhythm (but the difference was not replicated in Experiment 2).

During the test phase, subjects exhibited a bias to respond yes rather than no, and yes responses were considerably faster than no responses. Both groups of subjects generally (mis)classified new melodies as familiar. However, subjects trained on the “23..891/2R” rhythm were much better at rejecting different melodies as unfamiliar compared to subjects trained on the “12..89/1R” rhythm. Indeed, compared to their performance on new melodies, subjects trained on the “12..89/1R” rhythm were just as likely to misclassify different melodies as familiar, and do it just as quickly. In contrast, compared to their performance on new melodies, subjects trained on the “23..891/2R” rhythm were much less likely to misclassify different melodies as familiar, and when they did, responses were slow.

The observed differences between the two rhythm groups suggest that the “23..891/2R” rhythm is easier to learn than the “12..89/1R” rhythm, at least for English-speaking adults. An alternative which must be considered is that the “23..891/2R” rhythm is more compatible than the “12..89/1R” rhythm with the stress pattern of English. Is it possible that subjects learned nothing at all from the training melodies, but merely classified test melodies in accordance with their native language? Although this explanation might account for subjects’ treatment of different melodies compared to new melodies, it does not account for the treatment of old melodies, which enjoyed a recognition advantage over new melodies in both rhythm conditions. This was reflected in the proportion of melodies accepted as familiar, and also in the pattern of RTs. In particular, old nonkey melodies received the highest rates of familiar classifications and the fastest reaction times. Since the same melodies were randomly assigned as old nonkey melodies or new melodies to different subjects, the recognition advantage for old melodies in general, and nonkey melodies in particular, indicates that subjects learned something about their own training melodies, and that this knowledge influenced their responses to test melodies. Subjects did not simply classify test melodies by exploiting knowledge of their native language.

Among the old melodies, prototypicality effects were observed, in that the key melodies delineating category boundaries were at a disadvantage compared to nonkey
melodies. Key melodies were more likely than nonkey melodies to be misclassified as unfamiliar. When they were classified correctly, key melodies elicited slower responses, compared to nonkey melodies. Since the distinction between key and nonkey melodies was based on the nonmetrical constraints model of stress, the confirmation of the predicted prototypicality effects suggests that this model captures important aspects of the relevant mental representations. Like the other linguistic theories of stress, the nonmetrical constraints model assumes domain-specific knowledge, encompassing syllable structure, left/right directionality, and edge effects. Any theory of stress emphasizing these dimensions of variability would likely result in a similar set of key melodies.

**EXPERIMENT 2**

Adults do not necessarily approach learning a new stress pattern starting from the initial state of the language acquisition device. Jakobson, Fant, and Halle (1952, pp. 10–11) suggest that the perception of rhythm may be influenced by the rhythms of one’s native language. No support for this hypothesis was found by Bell (1977) in a controlled comparison between English (variable word stress), Bengali (initial phrasal stress), Polish (penultimate word stress), French (final phrasal stress), and Persian (final word stress). Nevertheless, the influence of one’s native language on the perception and processing of new linguistic rhythms remains unclear.

The subjects in Experiment 1 were native speakers of English. In order to assess the potential influence of subject’s existing knowledge of stress, a second experiment was conducted using native speakers of Brazilian Portuguese, recruited from around Oxford (these participants were primarily female, aged 20–40). English and Portuguese both assign stress within a few syllables from the end of a word, and both languages prefer nonfinal stress. However, while English prefers antepenultimate stress unless the penultimate syllable is heavy (in nouns; Chomsky & Halle, 1968), Portuguese favors penultimate stress (Bisol, 1992). If the results of Experiment 1 reflect an English-specific preference for nonfinal stress, then Portuguese speakers should show a similar, but weaker preference for nonfinal stress, because Portuguese generally places stress closer to the end of words compared to English.

In Experiment 2 we set out to test this hypothesis by giving Portuguese speakers a similar task to the one given to English speakers in Experiment 1. However, we doubled the length of the test phase, in order to test whether more responses could be elicited from each subject without affecting the pattern of results. The first block of 24 test melodies was generated exactly as in Experiment 1 to facilitate comparisons between the two experiments. These were followed by a second block of 24 additional test melodies. The two studies were otherwise identical in design.

**Results and Discussion**

The results for Experiment 2 generally replicated those of Experiment 1. The data from one subject were excluded from analysis on the basis of this subject’s training performance, which appeared to be at chance. During training, there was no significant difference in RT between training groups, $F(1,21)=2.2, p=.15$. No differences were found between the first and second sub-blocks in the longer test phase of Experiment 2. Subsequent analysis proceeded as in Experiment 1, but included both test blocks.
Results for the test phase of Experiment 2 are summarized in Figure 4, and results of the most specific statistical tests are summarized in the lower half of Table 3. In the main analysis, old melodies enjoyed a recognition advantage over new ones, but this memorization effect did not reach significance. Subjects trained on the “23..891/2R” rhythm generalized their training rhythm to novel melodies, but subjects trained on the “12..89/1R” rhythm did not. Results for the proportion of yes responses will be discussed first, followed by reaction time results. These will in turn be followed by analysis of prototype differences among old melodies. Finally, the influence of native language stress patterns is assessed by comparing the responses of Portuguese speakers (Experiment 2) and English speakers (Experiment 1), contrasting melodies with final versus nonfinal stress.
Yes/No analysis. Mean rates of yes responses are shown in Figure 4a. For both training groups there was a slightly higher rate of yes responses for old melodies compared to new melodies. Different melodies received a lower rate of yes responses compared to new melodies, but this difference was very small for subjects trained on the “12..89/1R” rhythm. A 2×3 ANOVA revealed significant main effects of familiarity and an interaction between familiarity and training rhythm, $F(2,42)=13, p<.001$ for the main effect, $F(2,42)=5, p=.01$ for the interaction. There was no main effect of training rhythm, $F<1$.

To test whether the familiarity effects were due to memory for old melodies or differentiation between new and different melodies, memorization and generalization contrasts were tested in 2×2 ANOVAs with training rhythm as the other factor. In the memorization ANOVA, the trend toward a memorization effect was not significant, $F(1,21)=1, p=.31$. There was also no main effect of training rhythm nor an interaction, $F(1,21)=2, p=.16$ for the main effect, $F<1$ for the interaction. In the generalization ANOVA, the main effect of generalization and the interaction with training rhythm were significant, $F(1,21)=14, p=.001$ for generalization, $F(1,21)=8, p=.01$ for the interaction. There was no main effect of training rhythm, $F<1$.

To verify that the generalization effects were due to responses of the “23..89/2R” group, separate t-tests were performed for each training group, comparing responses to new and different melodies. Subjects in the “23..89/2R” group generalized their training rhythm, accepting new melodies as familiar significantly more often than different melodies, $t(10)=4, p=.003$. Subjects in the “12..89/1R” group did not generalize their training rhythm, but accepted different melodies as familiar just as often as new melodies, $t(11)=0.8, p=.46$.

Reaction time analysis. Average reaction times for yes responses are shown in Figure 4b (again, reaction times for no responses revealed no clear differences for either training group). The pattern of reaction times is different for the two training groups. For subjects trained on the “23..89/2R” rhythm, yes responses were slower for different melodies compared to old or new melodies, which were about the same. For subjects trained on the “12..89/1R” rhythm, yes responses were fastest for different melodies, slowest for old melodies, and intermediate for new melodies.

Reaction times for each training group were tested for simple effects of familiarity. There were no significant RT effects in either training group for no responses. For the yes responses of subjects trained on the “23..89/2R” rhythm, a test of the null hypothesis that all three levels of familiarity were the same was significant, $\chi^2(2)=9.5, p=.01$, Friedman rank test for related samples. For subjects trained on “12..89/1R,” the effect of familiarity on RTs of yes responses showed a strong, but nonsignificant trend in the direction opposite to that predicted, $\chi^2(2)=5.1, p=.08$.

In order to test whether the familiarity effect observed for yes responses of subjects in the “23..89/2R” group was due to memory for training melodies or generalization of the training rhythm, reaction times for old and different melodies were compared to new melodies. The memorization effect (old vs. new) was not significant, $Z=.89, p=.18$, Wilcoxon matched pairs signed ranks, 1-tailed. There was a significant generalization effect, with slower responses for different melodies compared to new melodies, $Z=2.7, p=.004$.

Finally, a comparison between the two training groups confirmed that the generalization effect (the difference between new RTs and different RTs) was significantly different between the two training groups, $W=73, Z=3.5, p=.0004$, Wilcoxon rank sum test.
Prototype effects. In tests for prototype effects, analysis of the proportion of yes responses revealed nonsignificant trends in the predicted direction, and analysis of reaction times revealed significant effects. Starting with the proportion of yes responses, old nonkey melodies were classified as familiar more often than key melodies were, but this difference was not significant. For subjects trained on the “23..89/1R” rhythm, the fraction of melodies classified as familiar was 86% and 77% for nonkey and key melodies, respectively. Of the 12 subjects, the number showing a greater bias for nonkey melodies, key melodies, or no difference, was six, three, and two, respectively, \( Z = 1.2, p = .11 \), one-sided Wilcoxon matched-pairs test. For subjects trained on the “12..89/1R” rhythm, the fraction of nonkey and key melodies classified as familiar was 81% and 69%. Of the 11 subjects, the number showing a greater bias for nonkey melodies, key melodies, or no difference, was five, four, and three, respectively, \( Z = 0.9, p = .19 \). Analysis of reaction times showed that yes responses were slower for key melodies than for nonkey melodies. This difference was significant for subjects trained on the “12..89/1R” rhythm, \( Z = 1.72, p = .04 \), 1-tailed Wilcoxon matched pairs signed ranks test. The RT disadvantage for key melodies just missed significance for subjects trained on the “23..89/1R” rhythm, \( Z = 1.6, p = .055 \).

Final versus Nonfinal Stress

Again in Experiment 2, there were interactions between level of familiarity and training rhythm, but the effects were smaller than those observed in Experiment 1, possibly because of a difference between English and Portuguese speakers in bias against final stress. In order to test whether English speakers were biased more strongly than Portuguese speakers against final stress, responses for Experiments 1 and 2 were pooled and recoded by stress position (final or nonfinal). Only matched pairs of new and different melodies which differed in having final or nonfinal stress were included in this analysis. For the “12..89/1R” rhythm, 28 pairs of melodies met this criterion and were included. For the “23..89/1R” rhythm, 11 such pairs of melodies were included. Results are summarized in Table 4 and Table 5. Analysis of the proportion of yes responses is presented below, followed by analysis of reaction times. These analyses show (1) a nonfinal bias for speakers of both languages, (2) interactions between nonfinal bias and training rhythm, and (3) differences between English and Portuguese speakers consistent with the stress patterns of the two languages.

Analysis of the proportion of yes responses shows that for all language and training groups, there were more yes responses for melodies with nonfinal stress compared to melodies with final stress (see Table 4). Responses were analyzed by computing a signed chi-square nonfinal bias score for each subject from the distribution of responses (yes or no) by stress position (final or nonfinal). Bias scores for English and Portuguese speakers in corresponding training groups were compared using Wilcoxon’s rank sum test. The nonfinal bias was stronger for English speakers than for Portuguese speakers, but only for subjects trained on the “23..89/2R” rhythm, \( W_{23..89/2R}(11,12) = 94, p = .02 \) (two-tailed). Comparisons were also performed between training groups within each language category. For English speakers the nonfinal bias was significantly stronger for the “23..89/2R” group than for the “12..89/1R” group, \( W(12,11) = 157, p = .008 \). For Portuguese speakers the trend was in the same direction, but was not significant, \( W(12,11) = 105, p = .08 \).

Reaction times were analyzed by computing median RTs for each subject for melodies
with final and nonfinal stress (see Table 5). Only subjects with RT scores for both final and nonfinal melodies were included in the analyses. The first analysis, described below, tested whether the RTs were different for final and nonfinal stress. Separate tests were performed for each combination of language and training rhythm, and for yes and no responses (it was not possible to fully evaluate English RTs for the “23..891/2R” rhythm since only one subject in this condition gave a yes response for any melody with final stress). These tests reveal only nonsignificant trends, which are consistent with nonfinal bias for English speakers but not for Portuguese speakers. Further analysis tested whether the RTs for English and Portuguese speakers were different. Separate tests were performed for each training rhythm, and for yes and no responses. These tests show a significantly greater nonfinal bias for English speakers compared to Portuguese speakers.

Beginning with the English speakers, the pattern of RTs is consistent with a nonfinal bias in that yes responses were slower for melodies with final stress compared to melodies with nonfinal stress, while no responses showed the opposite pattern. However, the sample sizes are small, and tests of the differences in RT between final and nonfinal stress for English speakers were not significant, $Z=0.5, p=.6$ for yes responses of the “12..89/1R” group, $Z=0.9, p=.4$ for no responses of the “12..89/1R” group, and $Z=1.4, p=.16$ for no responses of the “23..891/2R” group (Wilcoxon matched pairs signed ranks, two-sided). The pattern of RTs for Portuguese speakers is more ambiguous. Portuguese speakers in the “12..89/1R” group gave slower responses for melodies with final stress compared to nonfinal stress whether they were answering yes or no. However, the difference between responses to final and nonfinal melodies was much smaller for yes responses than for no responses.

### Table 4

**Percentage of yes responses for paired melodies contrasting final and nonfinal stress**

<table>
<thead>
<tr>
<th>Stress</th>
<th>English&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Portuguese&lt;sup&gt;a&lt;/sup&gt;</th>
<th>English&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Portuguese&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonfinal</td>
<td>71</td>
<td>72</td>
<td>67</td>
<td>83</td>
</tr>
<tr>
<td>Final</td>
<td>55</td>
<td>60</td>
<td>2</td>
<td>45</td>
</tr>
</tbody>
</table>

Notes: <sup>a</sup>$n=96$ at each stress position (final and nonfinal). <sup>b</sup>$n=43$. <sup>c</sup>$n=40$.

### Table 5

**Mean RT differences for melodies with final versus nonfinal stress (ms).** Positive values indicate longer RT for final than nonfinal melodies. SE and number of subjects are given in parentheses.

<table>
<thead>
<tr>
<th>Language</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>70 (117, $n=12$)</td>
<td>–67 (90, $n=9$)</td>
<td>n.a.</td>
<td>–511 (420, $n=8$)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>256 (405, $n=11$)</td>
<td>844 (476, $n=8$)</td>
<td>305 (173, $n=9$)</td>
<td>–979 (672, $n=5$)</td>
</tr>
</tbody>
</table>
responses, a pattern which is consistent with training on the “12..89/1R” rhythm which emphasizes final stress. For Portuguese speakers in the “23..891/2R” group, the pattern of RTs is consistent with a nonfinal bias, but this pattern is also consistent with their training rhythm. Again, tests of the differences in RT between final and nonfinal stress for Portuguese speakers were not significant, \( Z=0.3, p=.8 \) for yes responses of the “12..89/1R” group, \( Z=1.8, p=.07 \) for yes responses of the “23..891/2R” group, \( Z=1.7, p=.09 \) for no responses of the “12..89/1R” group, and \( Z=1.5, p=.14 \) for no responses of the “23..891/2R” group (Wilcoxon matched pairs signed ranks, two-sided).

In comparisons between English and Portuguese speakers, the difference in nonfinal bias between English and Portuguese speakers was significant only for no responses of subjects trained on the “12..89/1R” rhythm, \( W(12,11)=123, Z=6, p=.6 \) for yes responses of the “12..89/1R” groups, \( W(9,8)=94, Z=2, p=.03 \) for no responses of the “12..89/1R” groups, and \( W(8,5)=29, Z=1.0, p=.3 \) for no responses of the “23..891/2R” groups, (Wilcoxon rank sum test). The direction of the difference in nonfinal bias (for no responses of the “12..89/1R” groups) is consistent with a stronger nonfinal bias for English speakers compared to Portuguese speakers.

If subjects were responding solely on the basis of knowledge of their native language (English or Portuguese), we would not expect to see a difference between the two rhythm groups in likelihood of responding yes to melodies with final or nonfinal stress. While both English and Portuguese speakers might be expected to avoid final stress, in keeping with the predominant stress pattern of their own language, this nonfinal bias should apply equally well to all test melodies unless there is some additional influence of training on responses. Yet there clearly was a difference between the two rhythm groups, as seen in Table 4. Subjects trained on the “23..891/2R” rhythm show a strong nonfinal bias, whereas subjects trained on the “12..89/1R” rhythm do not. Together with the fact that old melodies exhibited a recognition advantage over new melodies, these results indicate that subjects learned something about the training rhythms and were not merely exploiting knowledge of their native language. Nevertheless, responses of both English and Portuguese speakers were affected by the position of stress, consistent with the dominance of nonfinal stress in both languages. Moreover, differences between English and Portuguese speakers are in accord with the stress patterns of the two languages. English speakers exhibited a significantly stronger bias against final stress compared to Portuguese speakers, as measured by the proportion of yes responses for subjects trained on the “23..891/2R” rhythm, and also by reaction times for no responses for subjects trained on the “12..89/1R” rhythm. Other differences between English and Portuguese were consistent with these results, though they did not reach statistical significance. These results suggest that subjects were influenced by their native language when learning a new prosodic rhythm.

**EXEMPLAR SIMILARITY**

A fundamental question in psycholinguistics and cognitive psychology is whether productive use of knowledge involves abstract rules or reference to memories of specific instances. In general, an account of language acquisition must explain why language learners extract the generalizations they do from the examples they are exposed to. In the case of prosody, we might ask, for example, why there are “1R,” “2R” and “3R” stress patterns, but no languages with a “4R” stress pattern, placing stress four syllables from the end of
a word. What is the nature of prior knowledge and biases which determines the range of cross-linguistic variation? In Experiments 1 and 2, subjects trained on the “23..891/2R” rhythm treated novel melodies differently depending on whether or not they conformed to the training rhythm. Melodies which did not conform were less likely to be misclassified as familiar, compared to melodies which conformed to the training rhythm. The knowledge demonstrated by these subjects is implicit, in that subjects reported no conscious awareness of any systematicity in the rhythms. Were subjects applying abstract rules extracted from the training set and classifying novel melodies on the basis of whether or not they could be generated from the rules? Or were subjects referring to memory traces of training items and classifying novel melodies on the basis of similarity to the training items? The linguistic theories of stress outlined above assume the language learner is equipped with a specialized device for extracting appropriate generalizations from samples of the ambient language. An alternative is that the learner stores samples, or parts of samples, in memory, and that generalizations emerge from the distribution of items in the relevant mental space.

The distinction between rule-extraction and generalization from memory traces has received considerable attention in learning and memory research in many studies using artificial grammars to generate stimuli (where we have used attested stress patterns). At least three types of accounts have been proposed for the knowledge subjects acquire in artificial grammar learning tasks (cf. Redington & Chater, 1996). Some accounts, like the linguistic theories of stress, maintain that subjects acquire knowledge of the grammatical structure underlying the training stimuli (e.g., Manza & Reber, 1994; Reber, 1969, 1990). In contrast, exemplar-based accounts propose that generalization is based on memory traces of individual training stimuli. Whole-item exemplar models assume that training items are memorized in their entirety, and that test items are classified according to their similarity to items in the training set (Brooks, 1978; Brooks & Vokey, 1991). Fragment-based accounts claim that each stimulus is perceived as an assembly of chunks of various sizes, which are stored in memory and subsequently used as the basis for generalization (Perruchet & Pactau, 1990; Servan-Schreiber & Anderson, 1990). In the case of tone sequences, a chunk would be a subsequence of consecutive tones. Redington and Chater (Redington & Chater, 1996; Redington, 1996) have shown that simple fragment-based models can account for a wide range of artificial grammar learning results, and argue that fragment learning theories should serve as a null hypothesis, to be ruled out before considering explanations based on abstract rule extraction. In contrast to a host of studies examined by Redington and Chater, the results of these experiments stand out because they are not readily accounted for by simple fragment-based models of similarity.

According to the competitive chunking hypothesis (Servan-Schreiber & Anderson, 1990), learning of complex sequences involves building up a network of hierarchical chunks. Subsequent classification of novel stimuli is based on the extent to which a test item can be parsed into familiar chunks. The chunk familiarity of test items might be measured in terms of novel chunks, which do not appear at all in the training items (Meulemans and Van der Linden, 1997; Redington & Chater, 1996). Alternatively, chunk familiarity might be determined by taking account of the frequency with which chunks occur during training (Knowlton & Squire, 1994; Meulemans and Van der Linden, 1997; Servan-Schreiber & Anderson, 1990). We will consider first whether the distinction between novel and non-novel chunks in test items is sufficient to account for the experimental findings.
If performance in the present study was based on the number of chunks shared between test and training items, test melodies which received the highest rates of recognition responses should be those sharing the greatest number of chunks with training melodies. A simple chunking account, then, predicts that the average number of chunks shared between different melodies and training melodies should be greater for the melodies of the “12..89/1R” condition compared to the “23..891/2R” condition. To investigate whether the results of the current study are consistent with the chunking hypothesis, the similarity of test items to training items was analyzed using the 14 chunking models discussed in Redington and Chater (1996). Each model involves a different combination of assumptions about chunk size and attention to the beginnings and ends of stimulus items. Chunks may be either bigrams or trigrams (which terms will be used instead of the more accurate, but less familiar, bitones and tritones). There may or may not be explicit boundary markers at the beginning or end of an item. Attention may be focused on the initial or final chunk of each item, or on all contiguous chunks throughout each item. Test items are analyzed by each model, comparing the chunks of a test item with all the chunks of all training items. If every chunk of a test item occurs at least once in the corresponding training set, the test item is accepted by the chunking model. If any chunk of a test item is not found in the training set, the test item is rejected. The 14 models are identified in Table 6, along with

Redington and Chater impose additional constraints on the models so that they accept 50% of test items and reject the other 50%. For simplicity, these constraints are side-stepped in the current analysis. The results of the analysis do not hinge on the omission of these constraints.

### TABLE 6

Performance of chunking models for “12..89/1R” and “23..891/2R” rhythms. For each training rhythm, columns show the percentage of different melodies accepted by various chunking models (out of 28 potential different melodies for the “12..89/1R” rhythm, and 26 potential melodies for the “23..891/2R” rhythm). Asterisks identify models which are arguably consistent with the pattern of results observed in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Chunking Model</th>
<th>12..89/1R</th>
<th>23..891/2R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bigrams only</td>
<td>40%</td>
<td>73%</td>
</tr>
<tr>
<td>2. Bigrams with start marker</td>
<td>40%</td>
<td>69%</td>
</tr>
<tr>
<td>3. Bigrams with end marker</td>
<td>18%</td>
<td>73%</td>
</tr>
<tr>
<td>4. Bigrams with start and end</td>
<td>18%</td>
<td>69%</td>
</tr>
<tr>
<td>5. Trigrams only</td>
<td>11%</td>
<td>50%</td>
</tr>
<tr>
<td>6. Trigrams with start marker</td>
<td>11%</td>
<td>50%</td>
</tr>
<tr>
<td>7. Trigrams with end marker</td>
<td>4%</td>
<td>15%</td>
</tr>
<tr>
<td>8. Trigrams with start and end</td>
<td>4%</td>
<td>15%</td>
</tr>
<tr>
<td>9. Initial bigram only</td>
<td>100%</td>
<td>96%*</td>
</tr>
<tr>
<td>10. Final bigram only</td>
<td>18%</td>
<td>46%</td>
</tr>
<tr>
<td>11. Initial and final bigrams only</td>
<td>18%</td>
<td>46%</td>
</tr>
<tr>
<td>12. Initial trigram only</td>
<td>100%</td>
<td>81%*</td>
</tr>
<tr>
<td>13. Final trigram only</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>14. Initial and final trigrams only</td>
<td>4%</td>
<td>0%</td>
</tr>
</tbody>
</table>

2 Redington and Chater impose additional constraints on the models so that they accept 50% of test items and reject the other 50%. For simplicity, these constraints are side-stepped in the current analysis. The results of the analysis do not hinge on the omission of these constraints.
the number of different melodies accepted by each model for the “12..89/1R” and “23..891/2R” conditions.

Most of the chunking models are clearly inconsistent with the results of Experiments 1 and 2, either because they predict that more different melodies will be accepted in the “23..891/2R” condition compared to the “12..89/1R” condition (models 1–8, 10–11), or because they predict that very few different melodies will be accepted in either training condition (models 13–14). Only the initial bigram and trigram models (9 and 12) predict that different melodies will be readily accepted in the “12..89/1R” condition, with fewer different melodies being accepted in the “23..891/2R” condition. This means the general pattern of results observed in the experiments is consistent with the hypothesis that responses were based only on the initial two or three tones of each test melody. To investigate this possibility in more detail, each test trial was coded according to whether its initial bigram was consistent or inconsistent with the initial bigrams of the training items for that subject. Test trials were also coded according to initial trigram consistency with training items. Excluding melodies which were inconsistent with training items, the proportions of yes responses given by each subject to new and different melodies whose initial bigram (trigram) was consistent with training were subjected to a 2×2 ANOVA, with training rhythm as a between-subjects variable and familiarity (new or different) as a within-subjects variable. Because test items which were not consistent with training items were excluded, these analyses tested whether there were interactions between training rhythm and familiarity judgments over and above any effects of fragment-based similarity. In the analysis of initial bigram consistency, the interaction between training rhythm and familiarity was significant both for English speakers, \(F(1,22)=33, p<.001\), and for Portuguese speakers, \(F(1,21)=6, p=.02\). In the analysis of initial trigram consistency, the interaction was significant for English speakers, \(F(1,22)=41, p<.001\). The trigram interaction missed significance for Portuguese speakers, \(F(1,21)=3.6, p=.07\). These results indicate that subjects were sensitive to aspects of the stimuli beyond simple bigram or trigram familiarity. In short, the results of Experiments 1 and 2 cannot be explained by the simple fragment-based similarity models proposed by Redington and Chater.

In Meulemans and Van der Linden (1997), subjects studied letter strings generated by a finite-state grammar, then judged the grammaticality of novel letter strings. Test items with novel chunks (bigrams and trigrams which did not appear in any training items) were less likely to be judged grammatical compared with test items consisting entirely of chunks which were present in some training item. If subjects in the present experiments classified test melodies on the basis of novel chunks, accepting melodies with familiar chunks and rejecting those with novel chunks, then the different melodies for the “23..891/2R” rhythm should have more novel chunks than the different melodies for the “12..89/1R” rhythm. In order to test whether chunk novelty might account for the differences observed between rhythms “12..89/1R” and “23..891/R,” we computed the average number of novel chunks per different melody, and the fraction of different melodies with at least one novel chunk. Because the beginnings and ends of strings are particularly salient (Knowlton & Squire, 1994; Reber, 1967), we follow Meulemans and Van der Linden in measuring the novelty of anchor chunks (considering bigrams and trigrams only in initial or final position of test and training items) as well as overall novelty across all chunks. These measures of chunk novelty for different test items are shown in Table 7. All four
TABLE 7

Chunk Novelty. For each training rhythm, columns show the fraction of chunks in different test items which do not occur in any training items, and the fraction of different items which contain any such novel chunk. The figures for any chunks are based on all bigrams and trigrams, the figures for anchor chunks are based on initial and final bigrams and trigrams only.

<table>
<thead>
<tr>
<th>Rhythm</th>
<th>Fraction of Novel Chunks</th>
<th>Items with Novel Chunks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Any chunks</td>
<td>Anchor Chunks</td>
</tr>
<tr>
<td>12..89/1R</td>
<td>0.21</td>
<td>0.38</td>
</tr>
<tr>
<td>23..891/2R</td>
<td>0.17</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Meulemans and Van der Linden (1997) also consider associative chunk strength, a similarity metric which takes account of the frequency with which test fragments occur in the training set. The associative chunk strength for a test item was the average frequency with which the chunks (bigrams and trigrams) of the test item appeared in the training set. In their study, associative chunk strength was a significant predictor of grammaticality judgments, at least under some conditions (when the training set included just a small subset of all possible grammatical items). In order to test whether associative chunk strength might account for the results of Experiments 1 and 2, we computed associative chunk strength for old key and nonkey test items, new test items, and different test items (Table 8). These associative chunk strength measures are averages of the expected chunk strength of each melody against all possible sets of training melodies for a given target rhythm. Computed across all chunks of test and training items, the associative strength of different items is lower for the “23..891/2R” rhythm than for the “12..89/1R” rhythm (4.0 vs. 5.1). This pattern of associative chunk strengths corresponds to the observation that subjects trained on the “23..891/2R” rhythm were less likely than their “12..89/1R” counterparts to misclassify different melodies as familiar. So the pattern of classification for different melodies might be explained by supposing judgments were based on associative chunk strength. The other chunk-based measures do not fare so well. If chunk strength is computed only for anchor chunks (initial and final bigrams and trigrams), then the strength of different items is slightly lower for the “12..89/1R” rhythm than for the “23..891/2R” rhythm. This is not compatible with the pattern of results observed.

Although associative chunk strength computed across all chunks is compatible with the pattern of results observed for different items, it is clear that the prototype effects observed in Experiments 1 and 2 is not accounted for by associative chunk strength. The strength of old key melodies is greater than the strength of old nonkey melodies. If familiarity judgments were based on associative chunk strength the old key melodies should be more likely than old nonkey melodies to be classified as familiar. However, precisely the opposite pattern was observed in the experiments. In sum, all these measures of similarity...
based on bigram and trigram chunks fail to account for the experimental results. Another possibility is that training items are memorized whole, and that novel items are categorized according to their overall similarity to the set of whole training items (Brooks, 1978; Brooks & Vokey, 1991). To test whether global similarity between test items and training items might account for the experimental results reported above, we computed average distances between test and target melodies. The distance between two melodies was computed by counting the number of corresponding tones which differed in length, plus the number of corresponding tones which differed in pitch. This yielded a distance metric between zero and 7 for each pair of melodies. Distance metrics were computed separately for the two training rhythms. For each key melody, an average distance from training melodies was computed by weighting the distance to each other key and nonkey target melody by the probability that it was included in a randomly chosen training set. For each nonkey target melody, two average distance scores were computed. The old nonkey distance was a weighted average of the distances to key melodies, the distance of zero to itself (since by definition an old test melody is also a training melody), and the distances to other nonkey melodies. The new distance was a weighted average of the distances to key melodies and the distances to other nonkey melodies, excluding the item itself. Computed this way, the distance to training items is necessarily greater for new melodies than for old nonkey melodies. Finally, for each different melody, an average distance from training melodies was computed as a weighted average of the distances to key melodies and the distances to nonkey target melodies.

Average whole-item distance metrics are shown in Table 9. The pattern of whole-item distances between test and training melodies is generally consistent with the experimental results. For both training rhythms, old melodies (both key and nonkey) are most similar to (least distant from) training melodies, on average. New melodies are less similar to training melodies. Interestingly, the distance between different melodies and training melodies is very different for the two training rhythms. For the “12..89/1R” rhythm, the average distance from training melodies is less for different melodies than for new melodies. For the “23..891/2R” rhythm, the average distance from training melodies is greater for different melodies than for new melodies. This pattern of distances mirrors the finding that subjects in the “23..891/2R” condition were likely to reject different melodies as unfamiliar, while subjects in the “12..89/1R” condition were likely to accept them as familiar. When classifying new or different melodies as familiar, the responses of subjects

### TABLE 8

**Associative Chunk Strength.** For each training rhythm, columns show the associative chunk strength (based on frequency of chunks in training items) for key, nonkey, new, and different test items. The figures for all chunks are based on all bigrams and trigrams, the figures for anchor chunks are based on initial and final bigrams and trigrams only.

<table>
<thead>
<tr>
<th>Rhythm</th>
<th>All Chunks</th>
<th></th>
<th></th>
<th></th>
<th>Anchor Chunks</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Key</td>
<td>Non-Key</td>
<td>New</td>
<td>Diff</td>
<td>Key</td>
<td>Non-Key</td>
<td>New</td>
<td>Diff</td>
</tr>
<tr>
<td>12..89/1R</td>
<td>10.6</td>
<td>6.8</td>
<td>6.5</td>
<td>5.1</td>
<td>4.2</td>
<td>3.5</td>
<td>3.1</td>
<td>1.8</td>
</tr>
<tr>
<td>23..891/2R</td>
<td>8.8</td>
<td>6.4</td>
<td>6.0</td>
<td>4.0</td>
<td>3.9</td>
<td>3.6</td>
<td>3.2</td>
<td>2.1</td>
</tr>
</tbody>
</table>
in the “23..891/2R” condition were faster for new than for different melodies. Subjects in the “12..89/1R” showed the opposite pattern of reaction times to new and different melodies, consistent with the whole-item similarity scores of new and different items. Within the old melodies, the distance from training melodies is slightly greater for key melodies than for nonkey melodies. Again, this is consistent with the finding that key melodies were less likely than nonkey melodies to be classified as familiar, and when they were classified as familiar, responses were slower for key melodies than for nonkey melodies.

As a more formal test, the item distance-from-training scores were subjected to a 2×3 ANOVA, with training rhythm and familiarity (old, new, different) as factors. Both main effects and the interaction were significant, $F(1,176) = 4.3, p < .05$ for main effect of training rhythm, $F(2,176) = 13, p < .001$ for main effect of familiarity, and $F(2,176) = 3.9, p < .05$ for the interaction. In pairwise comparisons within training rhythms, the distance from training melodies was significantly less for old melodies than for new melodies, $t(58) = 2.5, p < .05$ for “12..89/1R,” and $t(56) = 2.5, p < .05$ for “23..891/2R.” For the “12..89/1R” rhythm, the distance from training melodies was not significantly different for different melodies and new melodies, $t(58) = .8, p = .45$. For the “23..891/2R” rhythm, the distance was significantly greater for different melodies than for new melodies, $t(56) = 2.7, p < .01$. Owing to the small number of key melodies (4 for “12..89/1R,” 6 for “23..891/2R”), the differences between key and nonkey melodies were not significant in a 2×2 ANOVA involving rhythm and key/nonkey as factors, $F < 1$ for all main effects and interactions.

The results of this section suggest that judgments of familiarity might have been based on whole-item similarity between test and training melodies, but were not based on comparisons between test melodies and melody fragments (bigrams or trigrams) stored in memory. Possibly, subjects perceived melodies holistically, and did not perceive them as sequences and subsequences of individual tones. If subjects did analyze melodies into bigram and trigram chunks, these chunks did not form the basis of generalization performance in classifying novel melodies. On the other hand, judgments of familiarity could have been based on comparisons between test melodies and whole training melodies stored in memory.

### GENERAL DISCUSSION

The main findings of Experiments 1 and 2 are that (1) subjects trained on the “23..891/2R” rhythm classified novel melodies in accord with their training rhythm, while speakers trained on the “12..89/1R” rhythm did not, (2) English speakers were biased more strongly than

#### TABLE 9

Whole-Item Distance. For each training rhythm, columns show the average distance (and SE of the mean) from training items (based on the number of tones of different length or pitch) for old key, old nonkey, new, and different test items.

<table>
<thead>
<tr>
<th>Rhythm</th>
<th>Key</th>
<th>NonKey</th>
<th>New</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>12..89/1R</td>
<td>3.70 (0.25)</td>
<td>3.67 (0.06)</td>
<td>3.91 (0.07)</td>
<td>3.83 (0.07)</td>
</tr>
<tr>
<td>23..891/2R</td>
<td>3.70 (0.19)</td>
<td>3.67 (0.07)</td>
<td>3.91 (0.07)</td>
<td>4.15 (0.06)</td>
</tr>
</tbody>
</table>
Portuguese speakers against final stress, (3) training melodies were perceived as more familiar than novel melodies conforming to the same rhythm pattern, and (4) melodies which delineate boundaries between rhythm categories were treated as less prototypical category members than others. These results demonstrate that subjects can generalize from specific exemplars and apply knowledge of at least some complex prosodic rhythms to novel melodies, after just a few minutes’ exposure to the rhythms. The finding that the “23..891/2R” rhythm was easier to learn contradicts intuitive expectations of the relative complexity of these two rhythms, and also goes against the relative cross-linguistic frequency of stress patterns based on these two rhythms. Moreover, a complete account of these findings must also explain the interference from prior knowledge of native language prosody and the differences observed among members of the same rhythm category.

What cognitive mechanisms underlie the ability to generalize prosodic rhythms? We can consider several possibilities regarding the present findings. Perhaps the melodies used in this study were not processed as prosodic rhythms, but were perceived and stored in memory as purely nonlinguistic sequences of sound. In that case, the results would reflect nonlinguistic rhythm processing, but would not reveal anything at all about the comparative difficulty of different stress patterns. Certainly, implicit learning of tone melodies will need further calibration against similar tasks using speech stimuli to evaluate the generality of these results. However, there is little or no evidence as yet that perception and memory of linguistic rhythm differs fundamentally from nonlinguistic rhythm processing (though cf. Lehiste, 1977). Moreover, general principles of linguistic theories of stress correctly predicted that some melodies would be treated as more peripheral members of rhythm categories compared to other melodies. Most importantly, the differences observed between English speakers and Portuguese speakers in their treatment of melodies with final versus nonfinal stress point to interference from knowledge of native language prosodic patterns. Interactions between native language prosody and the rhythm patterns of the present studies are not explained if the experimental stimuli were perceived and processed by a separate nonlinguistic module. Rather, the results support the view that subjects acquired knowledge of the same sort they bring to bear in processing the rhythms of their own language. This finding implies that experimental evidence can be brought to bear on questions of relative markedness between various prosodic patterns. This evidence may be crucial in deciding among competing learning models for linguistic prosody.

A simple generalization-from-exemplars model based on the distance between test melodies and training melodies is consistent with the bulk of these findings. The whole-item distance metric considered here compares two melodies by counting the number of tones differing in length plus the number of tones differing in pitch. The average distances between test and training melodies mirrors the differences observed between subjects learning the “12..89/1R” and “23..891/2R” rhythms, and would account for the recognition advantage of training melodies over novel melodies, as well as the differences observed between apparently more and less prototypical category members. It is not clear whether whole-item distance could also account for the apparent influence of native language prosody, though it seems likely that some such distance metric (comparing test melodies with prosodic patterns in the native language lexicon) could do.

Given the simplicity of the whole-item distance account, should linguistic accounts based on more abstract knowledge representations be abandoned? It would be premature
Learning prosodic rhythms to suggest this as a serious proposal, for several reasons. The distance metric considered above is ad hoc, limited in scope (applying only to equal-length strings, for example), and does not begin to account for the prosodic patterns observed (and not observed) in different languages. As one example, consider a “1R” stress pattern which places stress on the final syllable of each word. With respect to all the linguistic theories of stress considered above, this pattern can be uniquely determined by a single key melody, $\text{XXXXo}$, which demonstrates stress on a final short syllable in a word full of long syllables in other positions. If this were the only training melody, then the melody $\text{oooo}$ (which is consistent with the “1R” stress pattern) is at distance 4 from the training melody, while the melody $\text{XXXXo}$ (which is not consistent with the “1R” rhythm) is at distance 2. The whole-item distance model in this case cannot correctly generalize the training rhythm to include $\text{oooo}$ without incorrectly including $\text{XXXXo}$. In general, the whole-item distance model predicts patterns of generalization very different to those described by linguistic theories of stress. Perhaps some other distance metric would produce a more promising exemplar-based account of cross-linguistic prosodic patterns. Any such model would adopt a set of psychological dimensions along which to represent and compare prosodic patterns. If these dimensions ultimately correspond with the dimensions proposed by linguistic theories of stress (e.g., syllable weight, left/right directionality, and edge effects), then such an exemplar model may turn out to be an implementation of the knowledge structures incorporated by the linguistic theories.

What of more specific theories of stress acquisition? These models are based on summary knowledge of prosodic patterns, abstracted during training, and stored in the form of rules, constraints, or aggregate frequency-based statistics. The perceptron model of stress acquisition (Gupta & Tourretzky, 1994) learns to associate syllable weight strings with stress patterns. This model achieves the “12..89/1R” pattern only with great difficulty, and the “23..891/2R” pattern not at all. The model might be biased in favor of the “23..891/2R” pattern by finding a set of connection weights which produced this stress pattern, and stipulating these weights as the initial state of the learning mechanism. Of course, the resulting learning mechanism would no longer be a general-purpose learning device, but one highly specialized for this particular task. In any event, Gupta and Tourretzky’s main concern was not to demonstrate that stress patterns could be learned by a simple general-purpose algorithm, but to argue that theories formulated at a high level of abstraction (like linguistic theories of stress), do not necessarily shed much light on the actual processing mechanisms and causal factors underlying the phenomena targeted by the theory. That aim of their simulations is largely unaffected by the ability or inability of their model to account for the relative learnability of various stress patterns.

In the optimality theory constraint demotion account of stress learning, Tesar (1997a, b) suggests the language learner approaches the task with at least some constraints already ranked relative to each other. In particular, Tesar suggests the initial constraint ranking Weight-to-Stress >> Parse-Syllable >> {NonFinal, All-Feet-Right, All-Feet-Left}. Starting from this initial ranking, the “12..89/1R” stress pattern will require at least one round of demotion to achieve the hierarchy Weight-to-Stress >> All-Feet-Right >> {Parse-Syllable, NonFinal, All-Feet-Left}. The “23..891/2R” stress pattern will require at least two rounds of constraint demotion to arrive at the hierarchy Weight-to-Stress >> NonFinal >> All-Feet-Right >> {Parse-Syllable, All-Feet-Left}. To reverse the relative difficulty of
learning these two stress patterns, it is necessary to begin with an initial constraint hierarchy in which NonFinal >> All-Feet-Right. As a minimal departure from Tesar’s proposal, the hierarchy Weight-to-Stress >> NonFinal >> Parse-Syllable >> {All-Feet-Right, All-Feet-Left} would require at least one round of constraint demotion to achieve either “12..89/1R” or “23..89/2R” stress. A more decisive bias in favor of the “23..89/2R” rhythm could be achieved by simply assuming the initial hierarchy was already compatible with the “23..89/2R” rhythm. In any event, changing the initial hierarchy to account for the finding that “23..89/2R” was easier to learn than “12..89/1R” will affect the learnability of all stress patterns. Without doing a large set of simulations, it is difficult to determine whether some attested stress patterns would be unlearnable under any particular initial constraint hierarchy.

For the nonmetrical constraints model (Bailey, 1995), learning a stress pattern involves setting real-valued parameters which determine the relative importance of directionality and edge bias constraints. If a learner starts with directionality and edge bias parameters set to zero, then the “23..89/2R” pattern should be harder than the “12..89/1R” pattern to learn because the “23..89/2R” pattern requires greater changes in the parameter values (Figure 2). The finding that “23..89/2R” was easier to learn than “12..89/1R” could be accommodated by supposing that the initial parameter values were closer to the “23..89/2R” parameter region than the “12..89/1R” region. The finding that “23..89/2R” was easier to learn than “12..89/1R” would also suggest that the smaller size of the parameter region for “23..89/2R” had little or no effect on learnability.

The parameter-setting model proposed by Nyberg (1990) considers several related sets of parameters simultaneously. Each set of parameters under consideration is weighted according to consistency with the data observed so far, and generalization is based on the set of parameters which is most consistent with the data. Nyberg assumes without argument that the extrametricality parameter is [off] by default. This makes the prediction that stress patterns without extrametricality (e.g., “12..89/1R”) should be easier to learn than corresponding stress patterns with extrametricality (e.g., “23..89/2R”). Simply changing the default value of the extrametricality parameter to [on] would enable Nyberg’s model to learn “23..89/2R” more quickly than “12..89/1R,” and thereby accommodate the main result of Experiments 1 and 2. This change would have implications for the relative difficulty of other stress patterns, predicting, for example, that penultimate “2R” stress would be easier to learn than final “1R” stress, etcetera.

Of the various models of stress acquisition, only the parameter-setting model proposed by Dresher and Kaye (1990) predicts the present results. This model hypothesizes universal principles of stress combined with 11 binary parameters which determine stress placement. According to the Dresher and Kaye model, the language learner examines the ambient language for cues, and adjusts parameter settings until the target stress pattern is acquired. Learning is monotonic, in that each binary parameter can either remain in its initial, unmarked setting, or can be switched to its nondefault, marked setting. Once set, a parameter cannot return to its initial setting. In this model, the default parameter setting for extrametricality is [on], and the learning device will set the parameter to [off] only after encountering a sufficient number of counterexamples to the default setting (counterexamples being words with final stress). Since the “12..89/1R” rhythm involves the marked [off] setting, this stress pattern will require more data to learn than the otherwise-equivalent “23..89/2R” rhythm which involves the default [on] setting of the extrametricality parameter.
Although the Dresher and Kaye model correctly predicts the relative difficulty our subjects had in learning these two rhythms, one aspect of the Dresher and Kaye model seems incompatible with the present results. The Dresher and Kaye model “assumes a latency period during which a learner stores input, without attempting an analysis” (Dresher & Kaye, 1990, p. 172). The purpose of such a latency period is to allow access to a sufficiently broad set of examples from which to extract the correct parameter settings. Otherwise, several different (and mutually exclusive) parameter settings might be compatible with any given small set of data, and the learner has no way of knowing which set of parameters is correct. Subjects in Experiments 1 and 2 were tested after a training period which included just 96 melodies. The familiarity advantage of training melodies over novel melodies suggests that melodies were indeed stored in memory, but the ability of subjects learning the “23..891/2R” rhythm to reject melodies which did not conform to their training rhythm suggests either that this rhythm was learned very quickly, or that generalization performance depended largely on subjects’ prior experience with their own language. If the “23..891/2R” rhythm was learned very quickly, then the latency period proposed by Dresher and Kaye is in doubt. One alternative, outlined by Dresher and Kaye, is that parameter settings are continually re-evaluated on the basis of past experience, without regard to the current parameter values. This learning strategy eliminates the latency period before generalization can appear. It also makes the interesting prediction that discontinuities may occur in the patterns of generalization, when exposure to a few additional training items may result in a radically different set of parameter values.

Finally, since both English and Portuguese typically exhibit nonfinal word stress, the failure of subjects to learn the “12..891/1R” rhythm — which emphasizes final stress — could reflect interference from language-specific stress patterns rather than biases of universal grammar. The fact that English speakers exhibited a stronger bias against final stress compared to Portuguese speakers suggests that language-specific stress patterns had at least some effect on the results. The linguistic models of stress learning discussed above are aimed primarily at first language acquisition, and do not directly account for the observed differences between English and Portuguese speakers in the present study. The learning mechanisms employed by the adults in these studies may differ from the mechanisms employed by infants acquiring a first language. To investigate this possibility, the implicit learning of prosodic rhythm by infants is currently being studied, using stimuli similar to those employed here.

Received: November 12, 1997; revised manuscript received: August 12, 1998; accepted: September 15, 1998
REFERENCES


