Individual variability in cue-weighting and lexical tone learning

Bharath Chandrasekaran, Padma D. Sampath, and Patrick C.M. Wong

Roselyn and Richard Pepper Department of Communication Sciences, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

(Received 21 July 2009; revised 12 May 2010; accepted 13 May 2010)

Speech sound patterns can be discerned using multiple acoustic cues. The relative weighting of these cues is known to be language-specific. Speech-sound training in adults induces changes in cue-weighting such that relevant acoustic cues are emphasized. In the current study, the extent to which individual variability in cue weighting contributes to differential success in learning to use foreign sound patterns was examined. Sixteen English-speaking adult participants underwent a sound-to-meaning training paradigm, during which they learned to incorporate Mandarin linguistic pitch contours into words. In addition to cognitive tests, measures of pitch pattern discrimination and identification were collected from all participants. Reaction time data from the discrimination task was subjected to 3-way multidimensional scaling to extract dimensions underlying tone perception. Two dimensions relating to pitch height and pitch direction were found to underlie non-native tone space. Good learners attended more to pitch direction relative to poor learners, before and after training. Training increased the ability to identify and label pitch direction. The results demonstrate that variability in the ability to successfully learn to use pitch in lexical contexts can be explained by pre-training differences in cue-weighting.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3445785]

PACS number(s): 43.71.Rt, 43.71.Hw, 43.66.Hg, 43.71.Qr [AJ] Pages: 456–465

I. INTRODUCTION

Segmental (e.g., consonants and vowels) and suprasegmental information (e.g., pitch, duration) in speech are conveyed by multiple acoustic cues. Listeners tend to use a number of these cues in disambiguating speech sound patterns. Previous studies have demonstrated cross-language differences in the acoustic cues that are used in discriminating speech sounds (Francis et al., 2008; Gandour, 1983; Iverson et al., 2003). A well-studied example is the perceptual difficulties faced by Japanese listeners distinguishing between English sounds /r/ and /l/ (Bradlow et al., 1999; Bradlow et al., 1997; Iverson and Kuhl, 1996; Iverson et al., 2003). Japanese speakers do not have distinct categories for /r/ and /l/ sounds. From a cue-weighting perspective, Japanese participants do not attend to the third formant (F3) that acoustically distinguishes /r/ and /l/. The perceptual space of Japanese participants, relative to English participants is warped such that the second formant (F2) that does not distinguish between /r/ and /l/ is the dimension that is emphasized (Iverson et al., 2003). On the other hand, English participants’ perceptual space emphasizes the dimension (F3) that distinguishes these two sounds.

Adult second language learners thus have considerable difficulty in acquiring sound patterns of a foreign language (e.g., Iverson et al., 2003). However, laboratory studies have demonstrated that auditory training can help overcome the difficulty of perceiving foreign sound patterns (Bradlow et al., 1999; Lively et al., 1993; Logan et al., 1991; Wang et al., 2003; Wong and Perrachione, 2007). Feature-based approaches such as attention-to-dimension models (Francis et al., 2008; Francis and Nusbaum, 2002) and the Generalized Context Model (Nosofsky, 1987) argue that training induces changes in selective attention to cues that are necessary to categorize the foreign speech-sound contrast. Changes in cue-weighting may result in an increase in emphasis on relevant dimensions (stretching), and/or reductions in weights of irrelevant dimensions (shrinking). Together, training leads to warping of weights of dimensions underlying the perceptual space, resulting in increased efficiency in perceiving non-native contrasts.

Multidimensional scaling (MDS) approaches have been used to explore cues/dimensions that individuals use to define a perceptual space (Chandrasekaran et al., 2007a; Francis et al., 2008; Gandour, 1983; Iverson et al., 2003). A few MDS techniques have the capability of exploring individual differences in weighting (Carroll and Arabie, 1980). Previous multidimensional scaling studies have focused on cross-language examinations, and have generally found that cue-weighting is highly language-specific (Chandrasekaran et al., 2007a; Gandour, 1983; Iverson et al., 2003). While cross-language differences in feature-weighting and effects of training on changes in dimension weights have been relatively well-studied, individual differences in feature weighting, and their effects on learning have received much less attention. This is despite the fact that there is large variability in the successful learning of phonetic contrasts (Golestani and Zatorre, 2009; Iverson et al., 2005; Wong and Perrachione, 2007).

In the current study we examine the hypothesis that individual differences in learning to use pitch patterns of a tone language in a lexical context are associated with variability in cue-weighting. Languages that exploit phonologically contrastive variations in pitch at the lexical level are called
FIG. 1. Average fundamental frequency contours of time-normalized Mandarin Chinese tonal stimuli used in the training experiment. T1, T2, T3, and T4 refer to the four Mandarin tonal categories. T1 is phonetically described as ‘high-level’; T2 as ‘low-rising’; T3 as ‘low-dipping’; and T4 a ‘high-falling’. In terms of pitch height, T1 (high) and T3 (low) are most different; in terms of pitch direction, T2 (rising) and T4 (falling) are most distinct.

tonal languages (Gandour, 1994; Yip, 2003). Mandarin Chinese, a tone language, uses four tones: ma1 ‘mother’ [T1], ma2 ‘hemp’ [T2], ma3 ‘horse’ [T3], ma4 ‘scold’ [T4] (see Fig. 1). Tones 1 to 4 can be described phonetically as high level, high rising, low falling rising, and high falling, respectively. Voice fundamental frequency (F0) levels provide the dominant cue for tone recognition (Xu, 1997); perceptual data on Mandarin suggest that F0 is an important cue for recognition of citation-form tones (Howie, 1976). Other acoustic cues include amplitude (Whalen and Xu, 1992) and duration (Fu et al., 1998). Previous MDS studies of tone perception have found crosslanguage differences in dimensions utilized in tone perception (Francis et al., 2008; Gandour, 1983). Speakers of a contour tone language tend to attend more to a dimension related to pitch direction, whereas non-tone speakers tend to place more emphasis on pitch height (Gandour, 1983). This correlates well with the role of pitch in the two language groups; in tone languages, changes in pitch direction, irrespective of talker F0 (i.e., male or female speaker) signifies a change in lexical content. On the other hand, in all languages, pitch height is important in signaling speaker-specific information (Gandour, 1983). A recent study examined categorical perception of pitch direction in native and non-native speakers using parametric variation of the direction dimension [from level (T1) to rising (T2)] (Xu et al., 2006). Native speakers tended to exhibit more categorical perception of pitch direction, relative to non-native participants. Studies examining preattentive tonal processing using a neural index of change-detection, called the mismatch negativity (MMN), have demonstrated superior representation of pitch contour/direction in native speakers of Mandarin, relative to speakers of a non-tonal language (Chandrasekaran et al., 2007b; Kaan et al., 2007). Taken together, a consistent pattern across these studies is that native speakers of Mandarin selectively attend more to pitch contour/direction, and non-native participants tend to attend less to this dimension.

In this study, we first investigate if auditory training engenders changes in the weighting of cues important for tone perception in non-tone language speakers. We then examine the extent to which differences in successful use of L2 sounds in lexical contexts are driven by individual differences in cue-weighting. Can differential cue-weighting explain why some learners are better at utilizing L2 sounds, whereas others have considerable difficulty? This question is clinically relevant because it has clear implications for designing training paradigms that are specifically tailored to individuals. Our hypothesis, consistent with feature-weighting theories, predicts that the ability to emphasize/use pitch direction (a cue that native speakers of a tone language emphasize) contributes to individual variability in the ability to successfully use Mandarin pitch patterns in words by non-native (English) speakers.

In the current study, we use a sound-to-meaning training paradigm similar to that used in a previous training study (Wong and Perrachione, 2007) to examine the extent to which English speakers learn to use non-native pitch patterns in a lexical context. The sound-to-meaning paradigm is
unique in that the training involves learning to use sound patterns in a lexical context, not unlike native language acquisition. Participants are not trained to perceive sound patterns per se. Rather, the main focus of training is encouraging participants to learn words. In order to successfully learn the words however, perceiving the differences between sound patterns is important (Wong and Perrachione, 2007). Consistent with Wong and Perrachione (2007), in the sound-to-meaning training paradigm used in the current study, we provide feedback during learning, another important component of training (McCandliss et al., 2002). In addition, we incorporate other aspects of training that are known to benefit learning, i.e., use of multiple-talkers (Lively et al., 1993), and the use of natural, native (not synthesized) speech sound contrasts (Logan et al., 1991). Further, all participants underwent a long training protocol, spread over nine non-overlapping sessions, ensuring that they reached a plateau in overall performance. Thus, the training program ensures that all participants reach their maximum potential for learning, and ultimately yield asymptomatic performance in the last few sessions. To examine individual differences in cue-weighting, we tested the ability of participants to identify and label pitch direction, an important cue in lexical tone perception before and after training. In addition, we used individual differences scaling (INDSCAL), a three-way multidimensional analysis approach that examines individual variability in defining perceptual space.

II. METHODS
A. Participants

Participants were 16 (11 females, 5 males) young native speakers of American English (age: mean=22.05; standard deviation=3.4) with no history of neurological or hearing deficits. All participants passed a hearing screening test (audiological thresholds <25 dB HL across octaves from 500–4000 Hz). Participants were undergraduate or graduate students at Northwestern University. None of the participants had prior exposure to a tonal language. A musical history questionnaire was administered on all participants to determine extent of musicianship, a factor known to enhance lexical tone perception (Wong and Perrachione, 2007). All participants in this study had fewer than 6 years of formal musical training (instrumental or vocal). Further, none of the participants started musical training before the age of 12 years. Previous studies have demonstrated superior processing of lexical tones in musicians (Chandrasekaran et al., 2009; Wong et al., 2007; Wong and Perrachione, 2007; Wong et al., 2007). Since the aim of the current experiment was to elucidate the mechanisms underlying lexical tone perception, amateur musicians were excluded from this experiment to avoid a potential confound of musical-training induced plasticity. All procedures used in the experiment were approved by the Internal Review Board at Northwestern University. Including the tests (pre and post) and training sessions, each participant underwent a total of 12 experimental sessions.

B. Training Stimuli

Training stimuli consisted of 6 English pseudowords with superimposed pitch patterns resembling Mandarin Tones 1 (level), 2 (rising), 3 (dipping) and 4 (falling). English pseudowords were chosen because the main aim of the experiment was to examine how non-native English speakers, for whom pitch is not phonetically contrastive, learn to use pitch lexically. Stimuli that contain native phonological patterns (i.e., with legal native phonotactic patterns) are easier to learn than those that use nonnative phonological patterns (Feldman and Healy, 1998). Additionally, post-vocalic consonants in the pseudowords were all voiceless and similar to those used in a previous study (Wong and Perrachione, 2007).

There were six sets of words with four minimal pitch contrasts (Tones 1–4) per set. All four words in a set were identical except for the superimposed pitch pattern. Eight native speakers of American English (4 M, 4 F) were asked to record these pseudowords with a steady high pitch in a sound-attenuated chamber via a SHURE SM58 microphone onto a DAT recorder. All stimuli were sampled at 44.1 kHz. Eight native speakers of Mandarin Chinese (4 M, 4 F) produced the syllable /mi/ with the four Mandarin tones: high-level (Tone 1), rising (Tone 2), dipping (Tone 3), and falling (Tone 4). Tokens with creaky voice quality, as judged by the first author, were excluded. Five native speakers of Mandarin were asked to judge each hybrid (pseudoword+tone) stimulus for naturalness of pitch contour. Stimuli that were correctly identified by tone (accuracy >95%) and judged natural by all five participants were selected as training stimuli. The base pitch contours from each talker are shown in Fig. 1.

For creating the hybrid stimuli, the duration of the pitch contour in the /mi/ syllable and the pseudoword were calculated. The duration of the voiced portion of the CV syllable was matched in duration to the voiced portion of the CVC syllable using Praat software. Pitch patterns extracted from Mandarin productions were gender-matched with the native English pseudowords. That is, male and female pitch patterns were superimposed on male and female pseudoword production respectively. Pitch–synchronous overlap and add (PSOLA) method implemented in the software Praat was used to superimpose the pitch contour of /mi/ onto the CVC syllable. Thus the 24 tokens per talker were created by superimposing Mandarin tones on the English base pseudowords. The same procedure was reported for the other talkers yielding a total of 96 tokens (4 talkers × 6 pseudowords × 4 tones).

C. Pitch direction identification stimuli and procedures

A non-lexical pitch direction identification test was conducted with each participant before and after sound-to-meaning training. For the construction of this test, two male and two female speakers of Mandarin Chinese each produced five Mandarin vowels /a/, /i/, /o/, /e/, and /y/ with Mandarin Tone 1 (level tone). These vowels were then resynthesized similar to the procedures described above with Mandarin Tones 1 (level), 2 (rising), and 4 (falling) imposed. The dip-
ping tone (T3) was not included in the pitch direction identification test, since in the citation form, T3 does not have a consistent pitch direction (pitch falls during the first half of the tone and rises during the second half, see Fig. 1). The inconsistent pitch pattern for this tone makes it difficult to label for non-native speakers.

None of the speakers who produced the stimuli for pitch direction identification produced any of the training stimuli. The set of pitch patterns for each talker was different, and their naturally produced Tone 1 was used as a reference to first guide the resynthesis of Tone 1 and subsequently Tones 2 and 4. Including three random repetitions, there were 180 trials (four speakers × five vowels × three tones × three repetitions). Subjects heard one syllable at a time and were visually presented two different pictures (i.e., =level, ↑ =rising, and ↓ =falling), one depicting the pitch pattern of the auditory stimulus. Subjects were asked to press the response button that corresponded to the pitch pattern of the vowel they heard. The order of the two pictures was counterbalanced across trials. For example, button A=↑ (shown on the left side of the screen) and button B=↓ (shown on the right side). Participants were familiarized with the task before proceeding with the actual experiment. Because this task was designed to test subjects’ pretraining pitch direction identification ability, they received no feedback regarding their performance on this task, either from the computer or the experimenter. This task is considered “nonlexical” because subjects were not asked to identify or access words but to only attend to the pitch patterns. For further details about the stimuli for this test, see Wong and Perrachione (2007).

D. Pitch pattern (tone) discrimination stimuli and procedures

Participants also underwent a non-lexical tone discrimination task during which they performed speeded discrimination judgments of Mandarin tonal pairs. Participants were instructed to indicate via a response box whether the tone pairs presented were “same” or “different” in pitch direction and rate. Participants were familiarized with the task before proceeding with the actual experiment. Because this task was designed to test subjects’ pretraining pitch direction identification ability, they received no feedback regarding their performance on this task, either from the computer or the experimenter. This task is considered “nonlexical” because subjects were not asked to identify or access words but to only attend to the pitch patterns. For further details about the stimuli for this test, see Wong and Perrachione (2007).

E. Cognitive tests and procedures

All participants were evaluated with subtests from the Woodcock-Johnson Tests of Cognitive Abilities (WJ-Cog; Woodcock, 1997). Previous studies have shown that auditory working memory and phonological ability are associated with language learning (Baddeley, 2003a, 2003b; Jarrold et al., 2009; Papagno et al., 1991). The WJ-Cog subtests used in the current study included Sound Blending, Numbers Reversed, and Auditory Working Memory tests. In the Sound Blending test, participants were asked to synthesize temporally separated syllables into words. The test progressed in difficulty by finer splitting of words into component parts. In the test of Numbers Reversed, participants heard strings of numbers that increased in length, and had to repeat them back in reverse order. String length started at 4 digits and increased to 8. In the test of Auditory Working Memory, participants heard a mixed list of numbers and words, and had to repeat back first the words, and then the numbers. The list increased from 2 items to 8 items. Together, the three subtests evaluated phonological awareness, working memory, and auditory working memory. The means and standard deviations from these subtests are reported in Table I.

F. Training procedures

Participants underwent nine training sessions. Nine training sessions were chosen based on pilot data (n=4) that showed plateau in learning scores over sessions 6–9. During each of the nine training sessions, participants learned to associate the image of an object with one of 24 words. The 24 words were divided into six groups of four stimuli each. In the training session, stimuli in each group were minimally
TABLE I. Cognitive and Behavioral Profile of Good Learners and Poor Learners. GL=good learners, PL=poor learners. Numbers outside parentheses refer to group mean. Numbers within parentheses refer to standard deviation. Significance value (p-value) obtained from two-sample t-tests.

<table>
<thead>
<tr>
<th>Background</th>
<th>GL</th>
<th>PL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>0.99</td>
</tr>
<tr>
<td>Mean Age</td>
<td>22.3 (1.4)</td>
<td>21.8 (2.3)</td>
<td>0.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral/Cognitive testing</th>
<th>GL</th>
<th>PL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch direction identification</td>
<td>0.77 (0.09)</td>
<td>0.64 (0.08)</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>Pitch discrimination accuracy</td>
<td>0.96 (0.03)</td>
<td>0.83 (1.8)</td>
<td>0.10</td>
</tr>
<tr>
<td>Sound blending</td>
<td>81.75 (15.03)</td>
<td>76.00 (12.17)</td>
<td>0.42</td>
</tr>
<tr>
<td>Auditory working memory</td>
<td>87.86 (7.94)</td>
<td>80.88 (19.84)</td>
<td>0.37</td>
</tr>
<tr>
<td>Numbers reversed</td>
<td>74.50 (30.5)</td>
<td>78.38 (17.82)</td>
<td>0.76</td>
</tr>
</tbody>
</table>

contrasted by lexical pitch, and were repeated four times per talker (2F, 2M). The stimulus repetitions were blocked by talker to promote learning. At the end of each group, subjects were quizzed over the words they had just learned. Subjects heard one of the four words and selected the correct image from among the four they had just learned. The images were those of common objects (e.g., telephone, cow, etc.). Within a group of minimal pairs, the objects were not from the same category (e.g., zebra and cow are not in the same group). During the quiz, feedback (correct/incorrect) was used to allow participants to recognize and correct their mistakes. At the end of each training session, a final test was conducted. In this test, subjects were randomly presented with the 24 trained words and were asked to identify each word by selecting the corresponding drawing out of 24 possible choices. This procedure was repeated for each of the four talkers. No feedback was given during the final test. Subjects were given as much time as needed to identify the words. Subjects repeated this final test at the end of each of the nine training sessions. Each session, including the training, practice quizzes, and final test, lasted about 30 min. Subjects received four to five training sessions per week, with no more than a two day gap between sessions, and no more than one training session per day. To examine variability in learning success, participants were divided into good and poor learners based on the word identification scores from the final session (9th session) of training.

G. INDSCAL analysis

Multidimensional scaling analyses were conducted on the reaction time data obtained from the tone discrimination task. The fundamental assumption for this procedure is that the perceptual distance between two auditory objects can be discerned from the time taken to discriminate the two sounds (Nosofsky, 1992). That is, reaction time is greater for discriminating stimuli that are closer in perceptual space. For the current application, we utilized a type of three-way multidimensional scaling analysis called Individual Differences Scaling (INDSCAL). There are two specific advantages to INDSCAL analyses that are particularly relevant. First, the output of the analyses results in a group stimulus space, which is the representation of the four tones in Euclidian space. The resulting group stimulus space can be interpreted as is, i.e., does not require reconfiguration for the purpose of interpretation (Carroll and Arabie, 1980; Carroll and Chang, 1970). This will allow us to examine the dimensions that underlie tone perception in non-native speakers. Second, the output allows us to examine the weighting pattern for each participant contributing to the group space. We can then relate these dimension weights to learning success in the sound-to-meaning training task.

For the INDSCAL analyses, the input consisted of 32 (16 participants × 2) (pretest, posttest) separate 4 (stimulus)tones × 4 (stimulus tone) symmetric data matrices. Each data matrix contained distance estimates, i.e., the normalized inverse of reaction time (1/RT) for each paired comparison (T1 vs. T2, T1 vs. T3, T2 vs. T3, T1 vs. T4, T2 vs. T4, T3 vs. T4) of the four tones (Hume and Johnson, 2001). INDSCAL (Carroll and Chang, 1970) analyses of these 16 dissimilarity matrices were performed at n (where n = 1, 2) dimensionalities in order to determine the appropriate number of dimensions underlying the distances among the three tones or objects in a perceptual space. The output consisted of two matrices, a 4(stimulus tones) × n dimensions matrix of coordinates of the three stimulus tones on n dimensions (represented visually in a ‘group stimulus space’, (see Fig. 4), and a 16 × n matrix of weights of each of the 16 individual subjects.

III. RESULTS

A. Organization

The major aim of the experiment was to examine the nature of individual variability in the ability to use pitch in a lexical context. We evaluated the word learning performance from 16 participants who underwent a nine-session sound-to-meaning training program. In the next few sections we will report individual differences in word learning (Fig. 2) and effects of sound-to-meaning training on pitch direction identification and pitch discrimination tasks (Fig. 3). We report results from the INDSCAL analyses in terms of tone space.
and dimensional weighting (Fig. 4). The results of the cognitive and behavioral tests for the two groups are displayed in Table I.

B. Word learning performance

Sixteen participants completed all nine training sessions. At the end of each training session, participants’ word identification performance was assessed through the final test. This score was used as an index of learning success. The learning curves from each of the sixteen participants are shown in Fig. 2.

As can be seen from this figure, there was a high degree of variability in individual performance. Good learners (GL, n = 8) were defined as participants whose word identification test in the final session was above the median (top half). In contrast, poor learners (PL, n = 8) were defined as participants whose word identification scores during the final session were below the median (bottom half). The two groups did not differ on any of the cognitive tests (see Table I). The use of the median as the cut-off for group membership ensured that all participants in the study were utilized in the group-level analysis. A 2 × 2 (Group × Training sessions) repeated measures analysis of variance (ANOVA) revealed a main effect of group [F(1,14) = 44.07, p < 0.0001] and training session, F(8,112) = 84.50, p < 0.0001, and a group by training session interaction effect [F(8,112) = 22.06, p < 0.0001]. Post-hoc t-tests revealed that only the good learners showed improvement in learning between the 4th and 5th sessions [t(7) = 3.25, p = 0.014], and between 5th and 6th sessions [t(7) = 3.56, p = 0.01]. In contrast, poor learners showed no significant improvement between the 4th and 5th session [t(7) = 0.229, p = 0.826], or the 5th and 6th session [t(7) = 2.06, p = 0.078]. In the first session, differences between good and poor learners already emerge [t(14) = 2.74, p = 0.02]. The differences between the two groups (GL > PL) become more pronounced in the later sessions (e.g., for session 8, [t(14) = 7.84, p < 0.0001]); session 7 [t(14) = 8.18, p < 0.0001]; session 6 [t(14) = 7.6, p < 0.0001].

C. Pitch direction identification test

The pitch direction identification test was conducted before and after training. A 2 × 2 repeated measures ANOVA (Group (Good vs. poor learners) × Training (pretest vs. posttest)) conducted on pitch direction identification scores revealed a significant effect of group [F(1,14) = 7.90, p = 0.014], a main effect of training [F(1,14) = 24.92, p < 0.0001], but no significant group by training interaction effect [F(1,14) = 0.047, p = 0.832]. Overall, good learners were more accurate in pitch direction identification. Participants showed more accurate pitch direction identification scores after the training sessions, suggesting an effect of training irrespective of group membership (Fig. 3).

D. Pitch discrimination Test

A 2-way ANOVA (Group × Training) was conducted on the d-prime measure of sensitivity. Results revealed a main effect of group [F(1,14) = 4.675, p = 0.048]. Results revealed no significant main effect of training [F(1,14) = 0.03, p = 0.91] or interaction effect [F(1,14) = 2.68, p = 0.12]. Overall, good learners showed greater discrimination sensitivity relative to poor learners (Fig. 3). A 2-way ANOVA (Group × Training) was conducted on the reaction time data obtained from the AX-discrimination task. Reaction time data were collected from correct responses only. Results revealed a main effect of group [F(1,14) = 4.746, p = 0.047], and a main effect of training [F(1,14) = 5.34, p = 0.037] but no significant interaction effect between group and training [F(1,14) = 0.002, p = 0.967]. Overall, good learners were significantly faster than poor learners (Fig. 3). Participants were significantly faster posttest relative to pretest (Fig. 3).

E. INDSCAL analyses of reaction time data

A comparison of the badness-of-fit of the INDSCAL solutions for one- and two-dimensional spaces indicated that two dimensions best describe the reaction time data. The cumulative percentage of variance-accounted-for (VAF) by the two-dimensional INDSCAL model was 83%, an 11% increase over the one-dimensional solution (72%). These results indicate that two dimensions effectively characterize the subjects’ behavioral discrimination of the four tones. The group stimulus space of the two-dimensional INDSCAL solution for all 16 subjects’ dissimilarities data is shown in Fig. 4. Based on the configuration of the group stimulus space, Dim 1 is interpretively labeled ‘height’, as measured by av-
average F0 of each tone (T1, T2, T3, and T4). As per this configuration, T1 (high-level) is positioned toward one end of the axis, with T3, the low-falling-rising pitch contour positioned toward the other end of the axis. T2 and T4 fall between T1 and T3 on this dimension, as can be seen in Fig. 4. Dim 2 is interpretively labeled ‘pitch direction’, as measured by overall slope, measured from pitch onset to offset. T4 has negative slope (falling pitch direction), while T2 has a positive slope (rising pitch direction). In contrast to Dim 1, as per this configuration, T2 and T4 are the end points, whereas T1 and T3 fall between the two end points. T3 has both falling and rising contours, but the overall slope is closer to T1 than either T2 or T4. These dimensions are consistent with those described in a number of tone perception studies (Francis et al., 2008; Gandour, 1978; Gandour, 1983). To validate this solution we calculated the acoustic dimensions of average fundamental frequency (corresponding to pitch height) and overall slope (corresponding to pitch direction) for each talker’s production from the training data set (Fig. 4). When the two dimensions are plotted in a two-dimensional acoustical space (Peng, 2006), the configuration is similar to the solution obtained from the inverse of the reaction time data (i.e., the perceptual space). Across all talkers, for dimension 1, T3 which has a lower fundamental frequency is placed at one end point. T1, which has a higher average fundamental frequency is placed at the other end point. For dimension 2, T2, which has a positive slope, is contrasted with T4, which has a negative slope.

Results of a two-way ANOVA (Group×Training) of the mean subject weights per group (good versus poor learners) and per training (pre versus post) was conducted on each of the two dimensions. For dimension 1 (pitch height), a two-way ANOVA revealed no main effect of group \(F(1,14) = 3.78, p=0.072\), or training \(F(1,14)=2.35, p=0.15\) or interaction effect \(F(1,14)=1.78, p=0.534\). For dimension 2 (pitch direction), a two-way ANOVA revealed a main effect of group \(F(1,14)=5.003, p=0.040\), a main effect of train-
erable variability in adult lexical tone learning level cue-weighting. Our data suggests that there is consid-
erable emphasis on the pitch direction dimension as a function of training. Overall, good learners placed stron-
ger emphasis on the pitch direction relative to poor learn-
ers before and after training (Fig. 4). In a pitch direction identification test, good learners showed a super-
ior ability to label pitch direction relative to poor learn-
ers (Fig. 3). Similar to the dimension weighting results from
INDSCAL analysis, these differences exist even before train-
ing. Sound-to-meaning training increased pitch direction identification accuracy for good and poor learners. Based on
these results we conclude that variance in lexical tone learning is associated with the ability to use or attend to a specific
cue, pitch direction. Thus, pre-training cue-weighting ap-
ppears to be important in determining training-related plastic-
ity.

Pitch direction as a cue is known to be important to native speakers of Mandarin. Psychophysical evidence for
the relevance of pitch direction in tone languages comes pri-
marily from crosslanguage multidimensional scaling studies of
perceptual dissimilarity ratings. Pitch height and direction/
slope are dimensions that are known to underlie tone percep-
tual space for both native and non-native speakers of tone
languages (Gandour, 1983; Gandour and Harshman, 1978).
The relative importance of these dimensions, however, cru-
cially depended on native language experience. Mandarin
speakers tended to weight pitch direction more than English
speakers; while native English speakers weighted the height
dimension more that the direction dimension. Such differ-
ences in perceptual saliency suggest that language experi-
ence modulates the listener’s attention to cues that are par-
ticularly relevant in the native language. In Pike’s dichotomy
(Pike, 1948), tone languages can be roughly sub-divided into
contour-tone and register-tone languages. In contour-tone
languages, tones are described based on the movement of
pitch. Mandarin Chinese is classified as a contour-tone lan-
guage because of the high degree of pitch movement in the
tone language. Therefore, it is unsurprising that pitch direction is
an important cue for discriminating Mandarin tones. An im-
portant finding from the current study is that there is vari-
ability in the amount of emphasis placed on pitch direction
dimension among native-English speakers. Individuals who
utilize this cue better are more successful in learning to use
pitch in a lexical context. It is also relevant to note that good
learners are not better in general pitch perception ability.
Unlike dimension 2 (pitch direction), examination of dimen-
sion 1 (pitch height) weights did not yield significant group
or training-related effects. Pitch height is an important cue
for speaker identity (Baumann and Belin, 2010). Since the
multi-talker sound-to-meaning training did not emphasize
pitch height, the lack of training-related differences in this
dimension is not surprising. The sound-to-meaning training
paradigm used in the current study did not involve direct
training on discriminating pitch direction. To be successful
in this paradigm, participants had to learn to incorporate the
four tone categories and associate these with words. Feed-
back during training was restricted to whether the participant
correctly or incorrectly identified the word. Successful learn-
ing not only necessitated learning tonal categories, but also
involved learning to use pitch patterns in lexical context. The
paradigm is closer to language learning than training on
acoustics per se. Hence, the focus was on word learning.
Despite this, we found that participants tended to improve in
the ability to judge pitch direction with sound-to-meaning
training, suggesting that pitch direction is indeed an impor-
tant cue in lexical tone perception.

These data are consistent with previous studies on per-
ception of Mandarin tones by adult English listeners which
suggest that auditory training can indeed improve tone iden-
tification (Wang et al., 1999). Using a multiple talker para-
digm, this study demonstrated that native English speakers
were able to achieve a significant increase in the accuracy of
tone perception after auditory training. Based on these
results, the authors argue that listening to multiple talkers, with
a wide range of F0, reduces the emphasis on pitch height,
and causes participants to pay more attention to the previ-
ously less-attended direction dimension. In the current study,
using multidimensional scaling, we demonstrate that this ac-
count is indeed correct. Participants tended to attend more to
the direction dimension during the post-test relative to the
pre-test, suggesting that training induced an increase in rel-
ance of this specific dimension. Our data lend support to
the feature weighting models of perceptual learning that
posit that learning new sound patterns involves warping the
perceptual space so that more emphasis is placed on dimen-
sions that are relevant (Francis et al., 2008; Francis and Nus-
baum, 2002; Nosofsky, 1987). In this study we found pre-
training individual differences in cue-weighting within a

IV. DISCUSSION

Previous studies on perceptual learning of speech-sound
contrasts have thus far focused on crosslanguage differences
in feature-weighting (Chandrasekaran et al., 2007a; Kaan et
al., 2007; Wayland and Guion, 2004), and on the effect of
training on the warping of perceptual space (Iverson et al.,
2003; Francis et al., 2008). In the current experiment we
examined the extent to which individual differences in lex-
ical tone learning can be attributed to variability in lower-
level cue-weighting. Our data suggests that there is consid-
erable variability in adult lexical tone learning (Fig. 2) and
cue-weighting even within native English speakers who have
had no prior exposure to a tone language. INDSCAL analy-

ses of the reaction time data obtained from the tone discrimi-
nation task in the current experiment revealed two important
dimensions underlying non-native tone perception (Fig. 4).
Based on the tonal configuration and consistent with previ-
ous studies of lexical tone perception, dimension 1 was in-
terpreted as pitch height, and dimension 2 as pitch direction.
Our results indicate that weighting of dimension 2 increased
as a function of training. Overall, good learners placed
stronger emphasis on the pitch direction dimension (DIM2), rela-
tive to poor learners before and after training (Fig. 4). In a
pitch direction identification test, good learners showed a super-
iior ability to label pitch direction relative to poor learn-
ers (Fig. 3). Similar to the dimension weighting results from
INDSCAL analysis, these differences exist even before train-
ing. Sound-to-meaning training increased pitch direction
identification accuracy for good and poor learners. Based on
these results we conclude that variance in lexical tone learning
is associated with the ability to use or attend to a specific
cue, pitch direction. Thus, pre-training cue-weighting ap-
ppears to be important in determining training-related plastic-
ity.

Pitch direction as a cue is known to be important to
native speakers of Mandarin. Psychophysical evidence for
the relevance of pitch direction in tone languages comes pri-
marily from crosslanguage multidimensional scaling studies of
perceptual dissimilarity ratings. Pitch height and direction/
slope are dimensions that are known to underlie tone percep-
tual space for both native and non-native speakers of tone
languages (Gandour, 1983; Gandour and Harshman, 1978).
The relative importance of these dimensions, however, cru-
cially depended on native language experience. Mandarin
speakers tended to weight pitch direction more than English
speakers; while native English speakers weighted the height
dimension more that the direction dimension. Such differ-
relatively homogenous group. There were no differences between good and poor learners on cognitive scores or musical ability (Table 1). Differences between the two groups were restricted to weighting of the lower-level perceptual dimension (pitch direction), and the ability to utilize this cue in identification.

In conclusion, we sought to understand the nature of individual variability in lexical tone learning. We examined the extent to which variability can be accounted for by individual differences in lower-level cue-weighting. Utilizing multidimensional scaling we determined that successful learners tend to attend more to pitch direction, an important cue in Mandarin tone perception. Moreover, learning success was strongly predicted by pre-training ability to identify pitch direction. Sound-to-meaning training improved pitch direction. Sound-to-meaning training improved pitch 

**ACKNOWLEDGMENTS**

Work supported by the National Science Foundation (Grant No. BCS-0719666) and the National Institutes of Health (Grant Nos. R01DC008333 and R21DC009652).


