Perception of English Intonation by English, Spanish, and Chinese Listeners *

Esther Grabe[1], **Burton S. Rosner**[1], **José E. García-Albea**[2], **and Xiaolin Zhou**[3]

[1] *University of Oxford, England*
[2] *Universitat Rovira i Virgili, Tarragona, Spain*
[3] *Peking University, Beijing*

Key words

frequency modulation

intonation

multidimensional scaling

perception

Abstract

Native language affects the perception of segmental phonetic structure, of stress, and of semantic and pragmatic effects of intonation. Similarly, native language might influence the perception of similarities and differences among intonation contours. To test this hypothesis, a cross-language experiment was conducted. An English utterance was resynthesized with seven falling and four rising intonation contours. English, Iberian Spanish, and Chinese listeners then rated each pair of nonidentical stimuli for degree of difference. Multidimensional scaling of the results supported the hypothesis. The three groups of listeners produced statistically different perceptual configurations for the falling contours. All groups, however, perceptually separated the falling from the rising contours. This result suggested that the perception of intonation begins with the activation of universal auditory mechanisms that process the direction of relatively slow frequency modulations. A second experiment therefore employed frequency-modulated sine waves that duplicated the fundamental frequency contours of the speech stimuli. New groups of English, Spanish, and Chinese subjects yielded no cross-language differences between the perceptual configurations for these nonspeech stimuli. The perception of similarities and differences among intonation contours calls upon universal auditory mechanisms whose output is molded by experience with one's native language.

1 Introduction

Native language affects the perception of segmental phonetic structure and of prominence, of stress, and of tone. Semantic and pragmatic influences of intonation also

* *Acknowledgments:* This work was supported in part by grant R000237145 from the Economic and Social Research Council, United Kingdom, by grant PB96-1021 from the Ministry of Education, Spain, and by research grants to Xiaolin Zhou from the Chinese Ministry of Education. We thank Andrew Slater, Antonio Masip Cabrera, Danke Xie, and Jie Zhuang for technical assistance. Bob Ladd offered several helpful suggestions. The late Nancy C. Waugh made numerous useful comments on earlier drafts of this paper.

Address for correspondence. Phonetics Laboratory, University of Oxford, 41 Wellington Square, Oxford OX1 2JF, United Kingdom; e-mail: <esther.grabe@phonetics.oxford.ac.uk>.

vary with native language. Cross-language effects on the perception of similarities and differences between intonation contours, however, have received no attention. To begin filling this gap, we compared the perception of English contours by English, Iberian Spanish, and Mandarin Chinese listeners.

We first review cross-linguistic studies on the perception of speech and analogous nonspeech signals, concentrating on data from adults. Our work is *not* aimed at issues in second-language acquisition. We therefore refer only to a few studies relevant to our purposes that compare native speakers and non-native second-language learners.

We begin with experiments on the effects of native language on the perception of segmental phonetic structure. Next, we summarize cross-language studies on the perception of prominence and stress. Finally, we treat the few cross-language studies on semantic and pragmatic effects of intonation. These last studies must be distinguished from those on perceptual comparisons of intonation contours, where all previous experiments have concerned only single languages. Cross-language effects on the perception of a common set of intonation patterns have gone unexamined.

1.1
Native language and the perception of segmental phonetic structure

Native language affects an adult listener's perceptual judgments of the segmental structure of speech. Polka (1995) reviewed the evidence for such an effect on the identification and discrimination of speech contrasts in adults. The effect develops during the first year of life (e.g., Best, McRoberts, & Sithole, 1988; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). Burnham (1986) and Werker and Tees (1992) have reviewed data on this early transition from language-general to language-specific perceptual sensitivity. In a related vein, Pisoni, Lively, and Logan (1994) and Flege (1995) have summarized findings on the effects of native language on perception during second language learning.

Cross-linguistic differences occur in the perception of stops, fricatives, liquids and vowels. Seminal studies by Abramson and Lisker (1970, 1973) and Lisker and Abramson (1970) investigated the identification and discrimination of stops by native speakers of Latin American Spanish, American English, and Thai. The results showed that perceptual phoneme boundaries for voice-onset time in stops are language-specific. Williams (1977) went on to investigate voice-onset time in speech perception by Spanish and English monolinguals and in Spanish-English bilinguals. Significant differences appeared between phoneme boundaries for Spanish and English monolinguals and for monolinguals and bilinguals.

Hume, Johnson, Seo, Tserdanelis, and Winters (1999) compared the perceptual salience of place of articulation of Korean stops for Korean and American English listeners. The subjects heard transition stimuli and burst stimuli isolated from stops and were asked to identify the stops. Korean listeners performed better than American English listeners on the isolated transition stimuli. Hume and Johnson (2001) concluded that Korean listeners pay more attention to transitions following release bursts because Korean, unlike English, has tense, lax and aspirated stops.

Flege and Hillenbrand (1986) studied the effect of linguistic experience on perception of the English /s/–/z/ contrast in word-final position. These investigators used

two groups of native speakers of different American English dialects and two groups each of native speakers of French, Swedish, and Finnish that differed in English-language experience. Flege and Hillenbrand found that in identifying English fricatives as /s/ or /z/, the non-native subjects employed cues known to contribute to the perception of phonetic contrasts in their native language.

Miyakawa, Strange, Verbrugge, Liberman, Jenkins, and Fujimora (1975) tested the discrimination of [r] and [l] by adult native speakers of Japanese and American English. They made synthetic speech tokens varying from /r/ (i.e., English 'r') to /l/ by changing the initial frequency of F3, with F1 and F2 held constant. They also tested their subjects with just the isolated F3 portion of the stimuli, which were perceived as nonspeech "chirps." American English listeners discriminated the speech stimuli far better than did the Japanese listeners. Both the Americans and the Japanese, however, showed very good discrimination of the chirp stimuli.

Bladon and Lindblom (1981) found an effect of native language (English vs. Swedish) on the perception of vowels, although this issue was not central to their study. Flege, Munro, and Fox (1994) studied the perception of English and Spanish vowels by native speakers of English and Spanish. The English and Spanish subjects' ratings were similar for vowel pairs that were distant in the F1-F2 space but not for vowel pairs that were adjacent in an F1-F2 space. These results support the existence of both universal and language-specific effects in speech perception.

Finally, Mielke (2001) investigated the perception of /h/ in various segmental phonetic environments in Turkish, using Turkish, Arabic, American English, and French listeners. Turkish and Arabic allow /h/ in many environments, English allows /h/ in prevocalic position, and French lacks /h/. Clear cross-linguistic differences emerged. For the Turkish and Arabic subjects, /h/ was more salient than for the English and French subjects, while it was least salient for the French.

In summary, native and non-native language affect the perception of segmental phonetic structure. Some phonetic properties of speech, however, such as distance in a vowel space, may exert a common effect on speakers of different languages.

1.2
Effects of native language on the perception of aspects of prosody

We now turn to cross-language studies on the perception of prosody. These experiments concern the perception of prominence, stress, and tone. The effect of prosody on comprehension is not at issue here.

Prominence and stress. In sequences of reiterant speech, Lehiste and Fox (1992) manipulated duration and amplitude independently (see also Fox & Lehiste, 1987; Fox & Lehiste, 1989). When asked which syllable in a sequence was the most prominent, Estonian and English listeners gave significantly different answers. English listeners were more sensitive to amplitude cues, whereas Estonian listeners were more sensitive to duration cues. When the experiment was repeated with reiterant nonspeech signals, a comparable effect of native language emerged. Lehiste and Fox suggested that Estonian listeners might be more sensitive to duration than English listeners because Estonian is quantity-sensitive. Alternatively, since the cross-linguistic difference was particularly

striking in intonation-phrase final position, it might reflect language-specific expectations about phrase-final lengthening. Finally, these authors suggested that the results from the nonspeech condition might indicate that speech experience influences general auditory perception.

Following up work by Hermes (1997) and by House, Hermes, and Beaugendre (1997), Beaugendre, House, and Hermes (2001) investigated the perception of accent location. In five-syllable reiterant utterances such as /mamamamama/, the latter authors systematically varied the timing of a rapidly rising or falling pitch movement. Dutch, French, and Swedish subjects were asked to indicate which of the two medial syllables they perceived as stressed. The Dutch instructions used the term "Klemtoon," which is familiar to linguistically naïve speakers of Dutch (Hermes, personal communication). Listeners from the three language groups gave significantly different responses. For rising pitch movements, an earlier alignment of the pitch movement in the first of the two medial syllables led French listeners to judge the second syllable as prominent. In contrast, Dutch and Swedish listeners required a later alignment of the pitch movement for such a judgment. Dutch and Swedish speakers gave the same judgments for falling pitch movements, but no clear results emerged from the French subjects.

Dupoux, Pallier, Sebastián-Gallés, and Mehler (1997) investigated the perception of stress in nonwords by French and Spanish listeners. French as opposed to Spanish participants had difficulties in distinguishing nonwords that differed only in the location of stress. On the basis of their results, Dupoux and colleagues argued that French participants are "deaf" to stress contrasts because French, unlike Spanish, does not have lexical stress. Subsequently, Peperkamp and Dupoux (2002) proposed a typology of stress-deafness. They tested their typology by comparing the perception of stress in adult speakers of each of several languages: French, Finnish, Hungarian, and Polish. Speakers of some languages yielded more robust "stress deafness" effects than did speakers of others. In short, a variety of data show that native language affects the perception of prominence and stress.

Tone. The effect of native language on the perception of tone is not as solidly established as its effect on the perception of segmental phonetic structure. Gottfried and Suiter (1997) studied the identification of tones and of vowels in "silent center syllables" in Mandarin Chinese. In these syllables, only the initial and final portions of the syllable are presented. One Mandarin Chinese group of listeners and one American English group of listeners, who were studying Mandarin, took part in the experiment. The results of this study were not easy to interpret. An effect of native language on the perception of tone seemed to occur in some experimental tasks but not in others.

Lee, Vakoch, and Wurm (1996) investigated the effects of native language on the perception of Cantonese and Mandarin tones by Cantonese, Mandarin, and English listeners. Subjects were asked to determine whether two tones that they heard were the same or different. Native speakers of Mandarin or of Cantonese were better at discriminating tones from their own languages than were non-native listeners.

In experiments on the processing of lexical tone in Cantonese by Cutler and Chen (1997), Cantonese and Dutch listeners made speeded-response, same-different judgments of word or nonword pairs. Some of the pairs differed only in tone. Essentially,

the Dutch listeners produced the same results as the native Cantonese listeners. Both groups were more likely to judge a pair as identical when its members differed in tone only. Cutler and Chen concluded that their findings revealed the effects of simple perceptual processing rather than of specific linguistic knowledge. The processing of lexical tone distinctions may be slower than that of segmental distinctions. Since the experiment required rapid responding, the Cantonese listeners showed no native speaker advantage.

Huang (2001) investigated the origin of a tone sandhi process in Mandarin Chinese. She gave a discrimination task to Mandarin Chinese and American English listeners. No major cross-linguistic differences emerged from the main statistical analysis. Minor differences appeared in post hoc comparisons of responses to different tone pairs.

1.3
Native language and semantic and pragmatic effects of intonation

Cross-language studies have produced some data on the influence of intonation on semantic and pragmatic judgments. In her research on second language acquisition, Cruz-Ferreira (1984) demonstrated an effect of native language on the semantic influences of British English and European Portugese intonation. Cruz-Ferreira presented 30 British English and 30 Iberian Portuguese listeners with pairs of sentences that differed only in intonation. All subjects could speak the non-native language. Native speakers of each language produced the sentences. Subjects were asked to match each sentence with meaning glosses provided by the experimenter. Non-native speakers often did not choose the same gloss as native speakers, yielding a clear effect of language.

A more recent study by Chen, Rietveld, and Gussenhoven (2001) investigated language-specific effects of pitch range on pragmatic judgments. Dutch and British English listeners rated stimuli in Dutch and British English, respectively, on the scales CONFIDENT versus NOT CONFIDENT and FRIENDLY versus NOT FRIENDLY. The stimuli were lexically equivalent and varied in pitch range and contour. In both languages, perceived confidence decreased and perceived friendliness increased as pitch range was raised. However, at identical pitch ranges, British English was perceived as more confident and more friendly than Dutch. Chen et al. argued that the observed difference was due to the generally observed differences in the standard pitch ranges of Dutch and British English. Their study provides evidence of both universal and language-specific components in the pragmatic effects of intonation. It also is one of a number of studies that support the idea of a universal frequency code in intonation (Ohala, 1983). Gussenhoven (2002) gives an overview of work on this frequency code.

1.4
The perception of intonation contours

Various studies have examined the perception or pragmatic effects of intonation contours or units entirely within a single language (Bartels & Kingston, 1994; Gussenhoven, 1984; 't Hart, Collier, & Cohen, 1990; Kohler, 1987a, 1987b; Nash Mulac, 1980; Pierrehumbert & Steele, 1989). The previously established cross-language effects on various aspects of speech perception suggest that listeners from various language backgrounds would also yield different perceived patterns of resemblances between members

of a given, limited set of English intonation contours. We found no work, however, on the perception of intonation across language groups. We therefore undertook an investigation of native-language effects on the perception of English intonation contours. Specifically, we asked whether English listeners and listeners from other language backgrounds would perceive similarities and differences in pairs of the English stimuli in the same way.

2 Experiment 1

Stimuli were produced by resynthesis of a single English utterance, using each of 11 different fundamental frequency contours. Native speakers of English, Iberian Spanish, and Mandarin Chinese were asked to rate the differences between pairs of these 11 stimuli. English and Spanish are stress-accent languages. In both languages, stressed syllables can be associated with a limited set of pitch movements (for the term stress-accent, cf. Beckman, 1986). Spanish differs from English, however, in the inventory of observed accent shapes (Sosa, 1999). Mandarin is a tone language. Although it has postlexical intonational variation, it does not manifest phrase-long intonational contours constructed from basic pitch movements (Ladd, 1996).

The ratings were initially analyzed with multidimensional scaling (MDS) and with cluster analysis (CA). Multidimensional scaling represents the stimuli as points in a space of the lowest possible dimensionality, while preserving the rank-order relationships between the ratings (Everitt & Dunn, 1991; Schiffman, Reynolds, & Young, 1981). The space may have two or more dimensions. Perceptually similar stimuli end up in nearby positions in the space. Perceptually dissimilar stimuli fall in different regions. Cluster analysis is a standard procedure for confirming visual impressions of stimulus interrelationships in the parent MDS configuration. It takes the interstimulus distances from MDS and constructs a dendrogram (Everitt & Dunn, 1991). Perceptually similar stimuli appear on the same branch of the dendrogram, while perceptually diverse stimuli appear on different branches.

Multidimensional scaling has been used in numerous studies of the perception of vowels (see Rosner & Pickering, 1994, for a review). Outside linguistics, MDS and CA have been widely used in areas where people have difficulty in articulating the nature of perceived similarities or differences, for instance, in the study of tastes or smells. Similarly, the relationships between intonation contour shapes and perception are not particularly well formulated. Analysts frequently disagree on the phonetic description and phonological classification of English (and other) intonation contours. There is still no single, generally accepted approach to the description of intonation. For instance, in a recent edited survey of the intonation systems of 20 languages (Hirst & Di Cristo, 1998), a different approach to intonation description was taken for almost every language examined. Multidimensional scaling seems quite appropriate, then, for studies on intonation. In fact, Gussenhoven (1984), Herman and McGory (1999), and Huang (2001) have employed MDS in previous work on prosody.

On our initial hypothesis, English, Spanish, and Chinese speakers should yield different MDS configurations and different CA dendrograms for the 11 English stimuli differing in intonation. The British School of intonation analysis (e.g., Cruttenden,

1997) has long divided intonation contours into those ending in rising and those ending in falling nuclear pitch movements. Both types of contours were included in our stimuli. We expected the British subjects to separate the stimuli into two corresponding groups in the MDS map. Insofar as the perception of intonation contours is specific to speakers of the particular language that employs those contours, the results for the English subjects should diverge in various ways from those for the Spanish and Chinese listeners.

2.1 *Method*

Stimuli. We used 11 different intonation contours that are well attested in Southern British English and discussed in widely accepted descriptions of English intonation (see Cruttenden, 1997; Gussenhoven, 1984; Kingdon, 1958; O'Connor & Arnold, 1973). The contours can be taken as relatively uncontroversial. Each contour involved two intonation phrases and contained a prenuclear and a nuclear accent (for terminology, see Cruttenden, 1997). The spoken material, *Melanie Maloney*, was the same for each contour, providing a sequence of first name and surname.

From a phonological point of view, the 11 contours could be grouped in different ways, depending on the analysis chosen. For our immediate purposes, we classified the contours into two sets, based on the direction of the final pitch movement in the two accentual domains in each intonation phrase (for accentual domains, cf. Gussenhoven, 1990). The pitch movement is either falling (HL) or rising (LH). The location of an accentual domain is determined by the position of accented syllables. In the utterance *Melanie Maloney*, the prenuclear domain begins with the first accented syllable (*Me*—in *Melanie*) and ends just before the second accented syllable—*lo*—in *Maloney*. The nuclear accentual domain begins with the second accented syllable—*lo*—and extends to the end of the utterance. Seven stimuli had prenuclear and nuclear accentual domains ending in a falling pitch movement and formed Group HL. The other four had prenuclear and nuclear accentual domains ending in a rising pitch movement and formed Group LH.

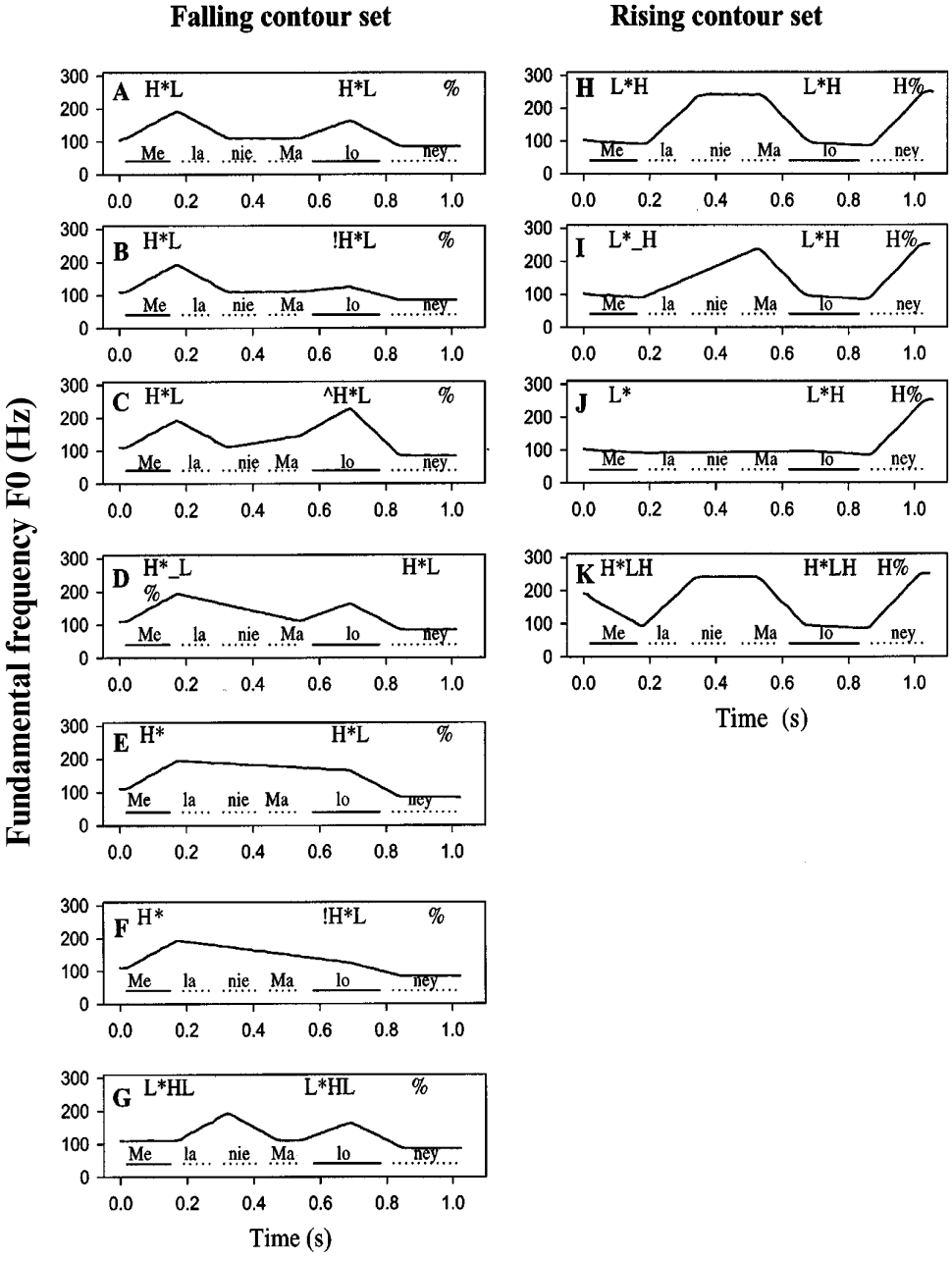
The imbalance in the number of HL and LH contours results from our separation of the contours into a rising and a falling group and from descriptions of Southern British English intonation. In Southern British English, phoneticians have described a wider variety of overall falling combinations of prenuclear and nuclear accents.

To make the stimuli, a male speaker of Standard British English initially produced *Melanie Maloney* with two different contours. One of these two utterances was produced as a statement in answer to the question, "What's Peter's girlfriend's name?." The resulting F0 contour contained two falling pitch accents, one on the first syllable *Me*—of *Melanie* and the other on the second syllable—*lo*—of *Maloney*. (These are the lexically stressed syllables.) The second utterance was produced in surprise to the announcement, "Peter's girlfriend's won the Nobel Prize!". Here, the resulting F0 contour contained two rising pitch accents. Both utterances were voiced throughout, so as to avoid any disruptions of the F0 contours.

The 11 experimental stimuli were generated from the two natural utterances. Seven contours based on the first utterance formed Group HL, and four contours based on the second utterance formed Group LH. Using PRAAT 3.8 (Boersma & Weenink,

Figure 1

Contours of F0(vertical axes) for the 11 resynthesized speech stimuli. Group HL (falling) and Group LH (rising) separated on basis of final pitch movement in each accentual domain. Horizontal lines beneath contours show locations of syllables in test utterance. Dotted lines represent unstressed syllables



1996) running on a PC computer, the stimuli were resynthesized using PSOLA (Carpentier & Moulines, 1990). PRAAT provides for manual shaping of an F0 contour before PSOLA resynthesis. Different starting utterances were used for the resynthesis of each group, since PSOLA can produce artifacts when the F0 contour for resynthesis departs widely from the starting contour. The two starting utterances were 1.033 and 1.054 s in duration for Groups HL and LH, respectively. This difference is well below the Weber faction for duration, which is at least 0.05; the difference in duration is therefore very unlikely to be perceptible. Each stimulus was saved to disc as a monophonic .WAV file sampled at a standard 44.4 kHz rate.

Figure 1 shows the F0 contours for Group HL on the left and those for Group LH on the right. The plots were obtained from the resynthesized stimuli. The intonation transcriptions follow the IViE (Intonational Variation in English) system for prosodic annotation (Grabe, 1998; Grabe, Nolan, & Farrar, 1998; Grabe, Post, & Nolan, 2001). The IViE system is an autosegmental-metrical, two-tone transcription system for intonation. It is based on the ToBI system (ToBI = Tones and Break Indices) derived from data on American English (Beckman & Ayers Elam, 1997; Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, & Hirschberg, 1992). The IViE system, however, rests on data on British English, the variety used for the present paper. Both systems assign prosodic labels such as H, L, * and % to accented syllables and intonation phrase boundaries. The Appendix gives corresponding ToBI transcriptions for Figure 1 and auditory descriptions that follow the terminology of the British School. The lines at the bottom of each plot in Figure 1 give approximate locations of the six syllables in each utterance. Solid and dotted lines represent accented and unaccented syllables, respectively.

Subjects and testing conditions. Three groups of paid volunteer subjects participated in the experiment. Subjects were always tested individually, and a PC controlled the experiment. The first group of listeners contained 22 native speakers of Southern British English, aged 19 to 25 years. The subject sat in a sound-attenuated experimental room in the Phonetics Laboratory of the University of Oxford. The visual display unit (VDU) in the room was attached via a splitter to the PC. In a separate control room housing the PC, the experimenter viewed a VDU that presented the same display as the subject's. A switch box activated a keyboard in the experimental room, so that the subject could send data to the PC. Output from the PC sound card was patched through to Sennheiser HD-320 earphones. The subject viewed a sheet with the experimental instructions as the experimenter read them aloud. All instructions and VDU messages were in English.

The second group of subjects contained 21 bilingual native speakers of Spanish and Catalan, aged 19 to 25 years. None could speak English, although most had some rudimentary knowledge of the language. The VDU messages and the instruction sheet were presented in Spanish. Subjects were tested in the Psycholinguistic Laboratory at the Universitat Rovira i Virgili, Tarragona, Spain. The subject sat in a sound-attenuated room with a VDU and keyboard connected to a PC. The experimenter sat in an adjacent control room that contained the PC. Stimuli were delivered over FoneStar FMC-690 earphones.

The third group included 20 native speakers of Mandarin Chinese, aged 17 to 23 years. All came from the Beijing area and could read some English. None spoke it fluently. All VDU messages and the instruction sheet appeared in standard Chinese

characters. The subject was tested in a sound-attenuated room at the Department of Psychology, Peking University. A VDU and keyboard were connected to a PC in the room itself. Output from the sound card was delivered over SONY MDR/CD470 earphones. The experimenter sat behind the subject, monitoring the latter's performance.

Procedure. Stimuli were delivered binaurally over earphones at a comfortable listening level of approximately 65 dBA. A program written in C controlled the experiment. At the start of a session, the subject heard each of the 11 stimuli individually. A message on the VDU informed the subject that a stimulus was available. When the subject pressed the ENTER key on the PC keyboard, the VDU screen went blank and a stimulus occurred after a 500 ms delay. The VDU message then reappeared. After the last familiarization stimulus, a new VDU message informed the subject that rating trials were about to begin.

Trials were self-paced. One pair of stimuli occurred on each trial. The pair was rated from 1 to 10 inclusive, depending on the degree of similarity or difference in the pitch movements of the two sounds (1 = very similar, 10 = very different). The subject was instructed that the members of a pair would never be identical and that a rating should only reflect the way in which the stimuli differed in pitch movement.

A VDU message to the subject indicated that pressing the ENTER key would initiate a rating trial. When the subject did so, the VDU went blank. A pair of stimuli was presented after a 500 ms delay; an interstimulus interval of 200 ms intervened between the end of the first stimulus and the start of the second. A second VDU message then instructed the subject to make a rating of 1 through 10, using the keyboard. The "0" key had been covered with a label of '10.' The subject could cancel the rating with the back-space key and enter a new number. After the subject made a rating and confirmed it by striking the ENTER key, the VDU went blank for 1000 ms. It then presented the message indicating the availability of a trial. An entire session took about 30 mins. A final VDU message informed the subject that the session had ended.

Each subject rated each of the 110 possible pairs of nonidentical stimuli just once. The 110 pairs occurred in a different random order for each subject. Ordered stimulus identifiers and the rating for each trial were written to a text data file.

Data analysis. The rating data were analyzed with a variant of INDSCAL (see Everitt & Dunn, 1991, pp. 84–88). This multidimensional scaling technique is implemented as *PROXSCAL* in the SPSS10.0 statistical package. We submitted the untreated, raw ratings (proximities) for each subject to a nonmetric analysis by *PROXSCAL*. Ashby, Maddox, and Lee (1994) demonstrated that averaging of data over subjects can produce misleading MDS results. The stimulus configuration in the derived space may fit the averaged data nicely but may not represent the results of any single subject. Therefore, an individual differences procedure such as *PROXSCAL* is the only safe MDS treatment of data from multiple observers. Data are not combined over subjects. *PROXSCAL* produces both the coordinates for the stimuli in the perceptual space and the weights that each subject gave to the axes of that space. The axes of the space may vary in relative importance across different subjects. The final stimulus locations in the perceptual space are optimal, given the set of weights for the subjects. We fitted both a two-dimensional and a three-dimensional solution to the data from each of the three groups of subjects.

To verify our visual impressions of the distribution of points in the MDS spaces, we undertook cluster analyses of the MDS interstimulus distances (see Everitt & Dunn, 1991; Chap. 6). Cluster analysis (CA) shows the stimuli that are perceptually closest and those that are perceptually farthest apart, whatever the dimensionality of the MDS space. We employed both single linkage (nearest neighbor) and complete linkage (furthest neighbor) algorithms, in order to be certain of the consistency of the analysis. The former locates perceptually nearby stimuli as quickly as possible. The latter forms different perceptual sets of stimuli as quickly as possible. The results in both cases are invariant over any monotone transform of the interstimulus distances. Therefore, the procedures pick out the main features of the MDS distances and are insensitive to details.

The number of subjects varied somewhat between groups. The MDS and CA procedures, however, do not depend upon the exact number of subjects in a group.

2.2 *Results*

Using a search procedure, PROXSCAL minimizes normalized raw stress. This measure is the sum of the squared differences between each MDS interstimulus distance and the corresponding proximity, after normalizing proximities across subjects to equate variances. Normalized raw stress decreases as a fit improves. The two-dimensional MDS solutions for the speech stimuli yielded normalized raw stress values of 0.0501, 0.0501, and 0.0451 for the English, Spanish, and Chinese subjects, respectively. Such small stress values indicate a good fit of the MDS solutions to the data. The corresponding values for the three-dimensional solutions were 0.0591, 0.0762, and 0.0684. The three-dimensional solutions obviously yielded no further improvement over the two-dimensional ones. The slight increases in normalized raw stress for the three-dimensional solutions probably represent entrapment of *PROXSCAL*'s search procedure in local minima (see Schiffman, Reynolds, & Young, 1981, p. 84 and p. 382).

MDS and CA findings. Figure 2 shows the two-dimensional MDS results from the three groups of listeners. Capital letters indicate the locations of the individual stimuli in the MDS space. The filled circles in each panel of Figure 2 represent the weights for the individual subjects. For clarity of graphical presentation, the weights actually obtained from *PROXSCAL* have been multiplied by five; otherwise, they would just appear as a smear not far from the origin. The dashed line in the first quadrant of each plot shows where equal weights for the two dimensions would fall. Points below that line represent subjects who gave more importance to the horizontal than to the vertical axis of the space. Points above the line indicate subjects who assigned relative importance to the axes in the opposite way. In each set of listeners, only one subject weighted the second dimension more heavily than the first. The configuration of the weight points in each plot indicates good agreement between listeners.

The English subjects separated the stimuli into Groups HL and LH, as predicted. The two stimulus groups occupied different regions of the MDS perceptual space. The Spanish listeners gave much the same results. With the apparent exception of stimulus G, the MDS configuration for Chinese subjects generally resembles those for the English and Spanish listeners. The latter subjects, who speak stress-accent languages, placed G

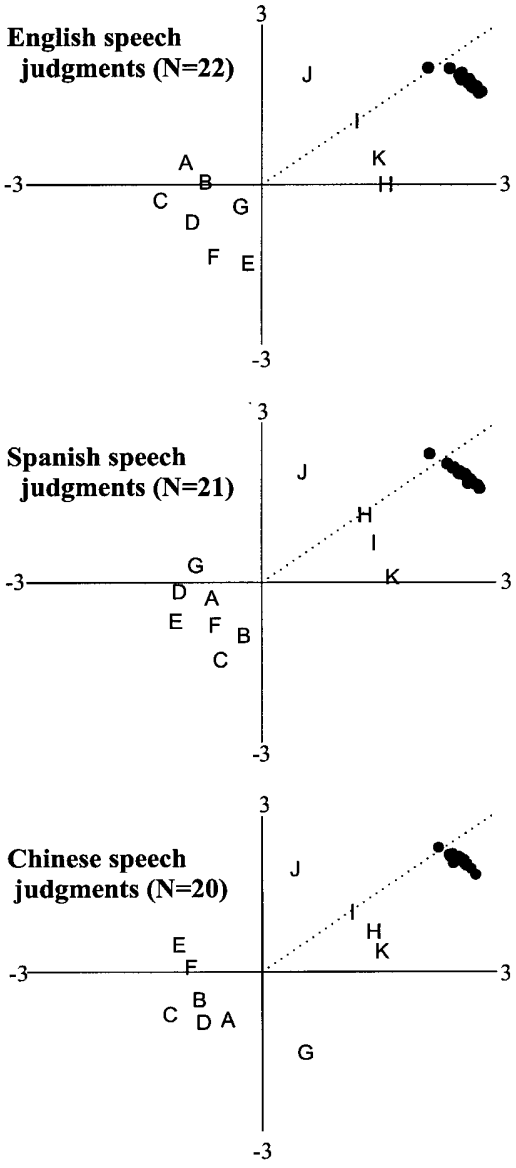


Figure 2

Two-dimensional MDS perceptual maps from ratings of speech stimuli by English, Spanish, and Chinese subjects. Letters correspond to contours in Figure. 1. Filled circles are five times subjects' derived weights for the two dimensions

together with the other contours in Group HL in the left half of the MDS space. The Chinese subjects, who speak a tone language, moved stimulus G into the right half of the MDS space. The dendrograms from the cluster analyses confirmed that all three groups of subjects separated Groups HL and LH. In particular, the cluster analyses for the Chinese subjects produced the same two subclusters as for the other subjects and clearly joined G with the other members of Group HL. No major cross-language differences emerged from the CA dendrograms.

Within stimulus Group HL, however, the MDS plots reveal some apparent perceptual differences between the three sets of subjects. For instance, the HL stimuli E

(H* H*L %) and F (H*!H*L %) are near neighbors in the English, the Spanish, and the Chinese MDS plots. All subjects judge stimuli E and F to be similar. Those stimuli, however, are at the top of the HL set in the MDS space for the Chinese subjects, at the bottom for the English subjects, and in the middle for the Spanish subjects. In contrast, the arrangement of stimuli within Group LH seems much the same for all three groups of subjects.

Two critical questions now arise about the MDS results for the speakers of the three languages. Are the apparent differences between the shapes of the HL stimulus configurations statistically reliable? And are the shapes of the configurations of the LH stimuli statistically indistinguishable? To answer such questions, we developed a statistical procedure, the configuration comparison test (CCT), based on correlation and regression analysis.

Configuration comparison test. The configuration comparison test takes three steps. It starts with the MDS distances, d_{ij}, between all pairs of points i and j, $i \neq j$, in an array of stimuli. An array of m stimuli yields $m(m-1)/2$ such distances. For example, the seven stimuli in Group HL yield 21 values of d_{ij}. All 11 stimuli in Groups HL and LH combined yield 55 values of d_{ij}.

Let $\{d_{ij\,\mathrm{V}}\}$ and $\{d_{ij\,\mathrm{W}}\}$ be the set of distances within a given stimulus array, for speakers of languages V and W, respectively. In the first step of the CCT, the Pearson product-moment correlation r_{VW} is computed between the two sets of distances. If two MDS configurations for a given set of stimuli are unrelated in shape, r_{VW} will be zero. If the configurations are at all similar, r_{VW} will be positive. Therefore, a 1-tailed test is used to determine whether r_{VW} is significantly greater than zero. If r_{VW} does not differ from zero, the configurations are unrelated in shape and the CCT terminates. No further useful information can be obtained.

If r_{VW} significantly exceeds zero, the CCT proceeds to its second step. Each set of distances is normalized to a mean of 1.0 and a standard deviation of 0.2. The normalized distances are denoted $d_{ij\,\mathrm{Vn}}$ and $d_{ij\,\mathrm{Wn}}$. Let M_{V} and s_{V} be the mean and standard deviation of $\{d_{ij\,\mathrm{V}}\}$. Then $d_{ij\,\mathrm{Vn}} = (d_{ij\,\mathrm{V}} - M_{\mathrm{V}})/5s_{\mathrm{V}} + 1$. A similar transform applies to $\{d_{ij\,\mathrm{W}}\}$. The transform compensates for differences between the arrays in overall spread and keeps each distance positive. It does not affect the shape or topology of each array of stimuli. It also leaves r_{VW} unchanged; indeed, the transform could be executed before the computation of r_{VW}.

In the third step of the CCT, the parameters A (the slope) and B (the intercept) are determined for the best-fitting straight line $d_{ij\,\mathrm{Wn}} = Ad_{ij\,\mathrm{Vn}} + B$, over all $m(m-1)/2$ pairs of distances. If the two MDS configurations are identical in shape, then A will be unity and B will be zero. If the two configurations are similar but not identical in shape, either or both of these conditions will not be met.

To test whether these conditions are satisfied, A and B are calculated along with their standard errors of estimate SE_A and SE_B, respectively. The standard errors lead to 2-tailed values of Student's t for A and B. Each t has $[m(m-1)/2-2]$ degrees of freedom. For the slope, which should be unity if the configurations are indistinguishable, $t_A = (1.0 - A)/SE_A$ is found. For the intercept, which should be zero if the configurations are indistinguishable, $t_B = B/SE_B$ is determined. If either value of t is

significant, the two configurations are reliably different, even though a significant r_{VW} indicates some similarity between them.

Since $\{d_{ij\,Vn}\}$ and $\{d_{ij\,Wn}\}$ have equal means and equal standard deviations, the values of A and B are identical for the regression of $d_{ij\,Vn}$ on $d_{ij\,Wn}$ and for the regression of $d_{ij\,Wn}$ on $d_{ij\,Vn}$. Furthermore, $A = r_{VW}$ and $B = 1 - r_{VW}$. (See Edwards, 1984, ch. 3, for details.) It is unfortunately impossible to test the hypothesis that r_{VW} does not differ from unity, by using Fisher's z transform. The value of z becomes infinite under that hypothesis.

In summary, given two MDS configurations, the CCT has three possible outcomes. First, if r_{VW} does not differ significantly from zero, the configurations are completely unrelated and different. Second, if r_{VW} significantly exceeds zero, and if A differs significantly from unity or B differs significantly from zero, the two configurations are similar but not statistically identical. Third, if r_{VW} significantly exceeds zero, and if A and B do not differ significantly from unity and zero respectively, the two configurations are statistically indistinguishable.

The outcome of the CCT is invariant under rotation or displacement of the axes of the MDS spaces. When applied to a given part of a configuration, it is invariant under any local rotation of that part. Finally, the CCT procedure applies to any pair of n-dimensional MDS spaces, $n \geq 2$.

To obtain values for r_{VW}, A, B, SE_A, and SE_B, we used SPSS 10.0 (SPSS, INC., 1999). As a further check, we also obtained Spearman rank-order correlations for comparison with r_{VW}. The two types of correlations always agreed in significance level. Due to the large number of tests reported in this paper, we have set α at a conservative value of .025 and have ignored differences that are significant at the .05 level.

As an example, Figure 3 shows $d_{ij\,Sn}$ for the Spanish speech results plotted against $d_{ij\,En}$ for the English results. The top, middle, and bottom panels display paired distances for the HL stimuli, the LH stimuli, and all stimuli combined, respectively. Regression lines and values of r_{ES} are shown; when appropriate, values of t_B and t_A are also given. A significance level accompanies any value. For Group HL, r_{ES} was not significant. Therefore, the two sets of subjects yielded reliably different configurations. For Group LH, r_{ES} again was not significant, apparently due to the different positions of stimulus H. There are fewer degrees of freedom for the LH than for the HL distances, however, so the LH finding should be treated cautiously. Finally, across all stimuli, the two configurations do not differ statistically. The correlation r_{SE} is significant; A and B do not differ significantly from zero and unity, respectively. This result basically reflects the fact that both the English and the Spanish subjects cleanly separate Groups HL and LH in the respective MDS spaces. It also shows that overall properties of two statistically indistinguishable configurations can mask important differences in detail.

Differences between language groups. We applied the configuration comparison test to the results for all 11 stimuli, for the HL stimuli (falls), and for the LH stimuli (rises), evaluating differences between the three possible pairs of language groups. Table 1 contains the findings on the speech stimuli. Each row shows the degrees of freedom for testing r_{VW} against zero and for each t-test. Then it displays the test results themselves and the CCT outcome.

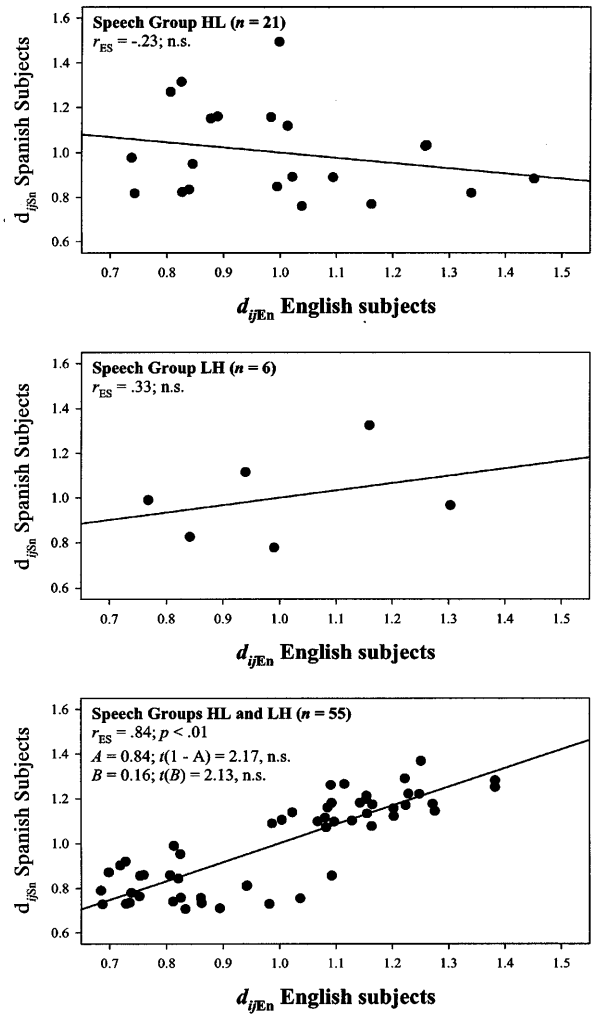


Figure 3

Standardized MDS interstimulus distances $d_{ij\mathrm{Sn}}$ for the Spanish results plotted against distances $d_{ij\mathrm{En}}$ for the English results. Top, middle, and bottom panels give findings on stimulus Group HL, Group LH, and both Groups combined, respectively. Each panel contains a regression line and relevant statistical output from the configuration comparison test (see text)

Using all stimuli, the shapes of the MDS configurations were statistically indistinguishable in the English/Spanish and the Spanish/Chinese comparisons (CCT in Table 1 marked + +). The correlations were significant, while A and B did not differ from unity and zero, respectively. This result arises from the numerous sizable distances connecting the stimuli in Group HL to those in Group LH (see Fig. 2). Although r_{SC} for the Spanish/Chinese comparison was significant, the slope A fell significantly below unity and the intercept B significantly exceeded zero. Therefore, the two configurations were similar but not identical (CCT in Table 1 marked +). This apparently happened because the Chinese listeners yielded a more tightly bound MDS configuration than did the Spanish listeners.

The configurations of the seven falling HL stimuli, however, differed reliably between all three languages (CCT in Table 1 marked 0). The correlations do not differ from zero for any of the three comparisons. Figure 4 shows regression plots for these comparisons.

TABLE 1

Two-way configuration comparison tests (CCTs) for English, Spanish, and Chinese subjects on English intonation contours embedded in speech stimuli. Results for all contours, falls only, and rises only. If correlation r_{VW} not different from 0, configurations unrelated; CCT result marked 0. Otherwise, if regression line slope $A \neq 1$ or intercept $B \neq 0$, configurations similar and CCT is +; if $A = 1$ and B = 0, configurations indistinguishable and CCT is ++

	$r_{VW} = 0$		$A = 1$		$B = 0$		
Test	*df*	r_{VW}	*A*	t_A	*B*	t_B	***CCT Result***
All stimuli							
English versus Spanish (A-K)	53	.84**	0.84	2.17	0.16	2.13	++
English versus Chinese (A-K)	53	.80**	0.80	2.45*	0.20	2.40*	+
Spanish versus Chinese (A-K)	53	.84**	0.84	2.13	0.16	2.25	++
Falls (Group HL)							
English versus Spanish (A-G)	19	–.23	–	–	–	–	0
English versus Chinese (A-G)	19	.25	–	–	–	–	0
Spanish versus Chinese (A-G)	19	.40	–	–	–	–	0
English versus Chinese (A-F)	13	.66**	0.60	1.63	0.34	1.63	0
Spanish versus Chinese (A-F)	13	–.08	–	–	–	–	0
Rises (Group LH)							
English versus Spanish (H-K)	4	.33	–	–	–	–	0
English versus Chinese (H-K)	4	.79	–	–	–	–	0
Spanish versus Chinese (H-K)	4	.73	–	–	–	–	0

** $p < .01$; * $p < .025$; – = test irrelevant

The Chinese subjects treated HL stimulus G quite differently from the other two groups. We eliminated distances for that stimulus within the HL group and ran new configuration comparison tests for the Chinese listeners against each of the other two sets of subjects. As Table 1 shows, the Spanish/Chinese results remained unchanged (CCT in Table 1 again marked 0). Although r_{EC} was only .66, the English/Chinese configurations now became statistically indistinguishable (CCT in Table 1 now marked ++). Both of these groups of listeners placed stimuli A, B, C, and D close together, with E and F in a more distant group, yielding the significant correlation (see Fig. 2). The Spanish subjects did not behave this way.

Finally, Table 1 shows that the configurations for the rising LH stimuli differed reliably in all comparisons (CCT in Table 1 marked 0). As noted above, the few degrees of freedom available for these tests may have prevented the correlations from reaching significance in at least the English/Chinese and the Spanish/Chinese comparisons.

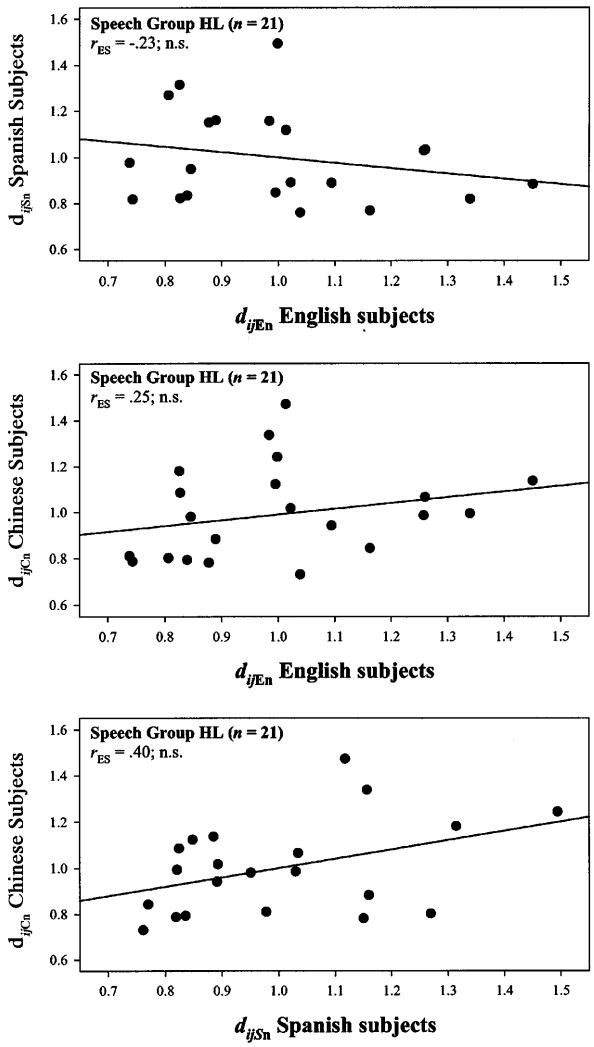


Figure 4

Standardized MDS inter-stimulus distance plots for group HL for the English/Spanish (top panel), English/Chinese (middle panel), and Spanish/Chinese (bottom panel) comparisons. Each panel contains a regression line and relevant statistical output from the configuration comparison test (see text)

2.3 *Discussion*

Both our two-dimensional MDS results and Gussenhoven's (1984) three-dimensional MDS configuration show that English listeners make a major perceptual split between HL and LH F0 contours. Our Spanish and the Chinese subjects also separated the stimuli into HL and LH groups. Further comparisons between Gussenhoven's results and ours cannot be made. Gussenhoven does not provide a two-dimensional solution, and the stimuli carrying his F0 contours were presented in five different carrier contexts. Our stimuli were presented without context.

Reliable differences, however, appeared in the perceptual organization of our stimuli within Group HL in all three cross-language comparisons. The configurations for this stimulus group differed markedly between the English and the Spanish subjects. We could not identify one or two stimuli that underlay most of this effect. The location

of rising-falling contour G in the Chinese MDS map differed noticeably from its position in the English and Spanish maps. Cluster analysis, however, still placed G in Group HL for the Chinese listeners. Configuration comparison tests revealed that stimulus G explained the difference in the perception of Group HL by the English as against the Chinese speakers. It did not do so for the Spanish/Chinese comparison.

Cross-language differences were found within each of the three comparisons for Group LH. Given the small number of contours within that group, this finding must be treated suspiciously. Experiments using a larger number of LH contours are obviously needed.

The consistent cross-language MDS split between HL and LH stimuli suggests that perception of intonation contours may start with the activation of universal auditory mechanisms that process relatively slow frequency modulations. These mechanisms would determine the basic cross-language perceptual distinction between Groups HL and LH. A listener's experience with her native language could modify the outputs of the basic auditory mechanisms, producing different language-dependent perceptual configurations for various subsets of stimuli.

To test this hypothesis, we made frequency-modulated (FM) pure tones whose contours duplicated those of the speech stimuli. British English, Spanish, and Chinese listeners were asked to rate the dissimilarity of pairs of these nonspeech stimuli. If the perception of F0 contours in speech draws initially on universal auditory mechanisms that are not speech-specific, all three MDS configurations for the nonspeech FM stimuli should be two-dimensional and should again show a clear division between Group HL and Group LH. Furthermore, the perceptual organization of stimuli within each of those two stimulus groups should not differ significantly across languages, in contrast to the results for speech.

3 Experiment 2

We collected similarity judgments using 11 frequency-modulated pure tones, presented in pairs. The frequency changes in the stimuli duplicated the F0 contours of the speech stimuli in Experiment 1. Again, we tested English, Iberian Spanish, and Mandarin Chinese subjects. None had participated in Experiment 1. This precaution avoided any possible cross-contamination between judgments of the speech and the nonspeech stimuli. We expected all three groups of subjects to segregate the HL and LH contours in two-dimensional MDS spaces and in addition to yield similar perceptual organizations of the stimuli within each group. Finally, within-language comparisons of MDS arrays for the speech and the corresponding FM stimuli would show how far the former configurations diverge from the latter.

3.1 *Method*

Stimuli. The 11 frequency-modulated (FM) sine waves were generated with Cool96 (Syntrillium Software Corporation, 1996). For each contour in Figure 1, we made a sinusoidal stimulus that contained identical linearly frequency-modulated segments. The segments were produced individually and then pasted together. Any resulting artifacts

were deleted. The FM patterns therefore were copies of the F0 patterns in the speech stimuli. Stimulus onsets and offsets were shaped with 20 ms linearly rising and falling amplitude modulations, respectively. Otherwise, stimulus amplitude was constant and equal across stimuli. Fundamental frequency tracks obtained with PRAAT showed that the contours of the FM stimuli duplicated those of the speech stimuli.

Subjects, testing conditions, procedure and data analysis. Three new groups of paid volunteer subjects were studied. The first contained 20 native speakers of Southern British English, aged 19 to 23 years. The second group contained 20 native speakers of Spanish and Catalan, aged 19 to 25 years. The third included 20 native speakers of Mandarin Chinese, aged 18 to 22 years. The testing conditions for each set of subjects were the same as those of Experiment 1, as were the procedure and data analysis.

3.2 *Results*

The two-dimensional MDS solutions for the sinusoidal FM stimuli had normalized raw stress values of 0.0505, 0.0464, and 0.1016 for the English, Spanish, and Chinese subjects, respectively. For the three-dimensional solutions, the corresponding values were 0.0671, 0.0839, and 0.0877; the first two values indicate entrapment in local minima. The two-dimensional solutions were obviously adequate.

Comparisons across speakers of different languages. Figure 5 (overleaf) shows the two-dimensional results for the FM stimuli. It is organized exactly like Figure 2. The letters in Figure 5 correspond to those in Figures 1 and 2, but they are now marked with primes to distinguish the FM stimuli from the speech stimuli. In each group of listeners, two subjects weighted the second dimension more heavily than the first. The weights for the English and the Chinese subjects seem slightly more divergent for the FM stimuli than for the speech stimuli. These results may reflect differences between subjects or differences between stimuli in Experiments 1 and 2.

The MDS plots confirmed our predictions. All three groups of subjects perceptually separated the FM stimuli in Group HL from those in Group LH. Furthermore, for all three groups of listeners, the perceptual configurations of stimuli within Groups HL and LH are now much the same. Within Group LH, however, J,. was consistently set quite far apart from its other three partners. Cluster analyses supported these findings.

Table 2 (overleaf) gives the results of configuration comparison tests for the FM stimuli, pitting the three sets of subjects pairwise against one another. Not one of the nine comparisons represented in the table revealed a reliable difference between MDS configurations. Each pair of configurations tested proved statistically indistinguishable.

Comparison of results for speech and for FM stimuli. Using the CCT, we evaluated differences in the MDS positions of speech and FM stimuli within subjects who are native speakers of the same language. Table 3 contains the findings. The results are separated as usual into tests for all stimuli, for the falling HL stimuli, and for the rising LH stimuli.

Tests over all contours showed that the speech and FM configurations were indistinguishable in shape for Chinese subjects. For English subjects and for Spanish subjects, however, the speech and FM configurations were similar but distinguishable. The

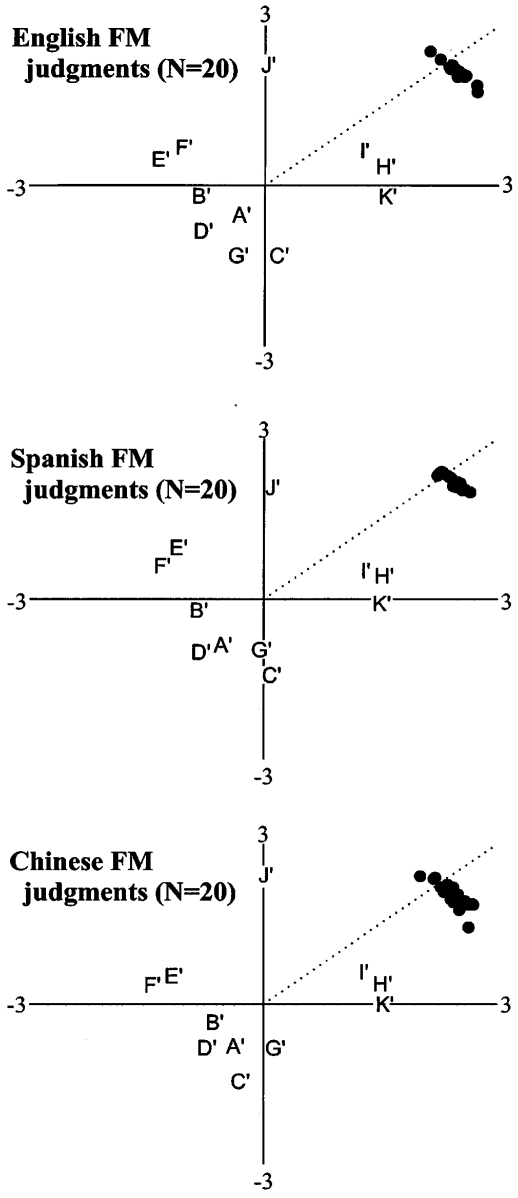


Figure 5

Two-dimensional MDS perceptual map for ratings of FM stimuli by English, Spanish, and Chinese subjects. Letters correspond to contours in Figure 1. Primes distinguish FM stimuli from speech stimuli of Experiment 1. Filled circles as in Figure 2

correlation was significant, but *A* fell significantly below unity and *B* exceeded zero. The FM configurations always seemed more scattered than the speech configurations.

The English and the Spanish results also showed reliable differences between the perceptual configurations for the Group HL falling speech and FM stimuli. In contrast, the Chinese findings were statistically indistinguishable. Finally, the rising contours produced a barely reliable difference between speech and FM stimuli only for the English results. The configurations for the speech and the FM rising stimuli were statistically indistinguishable for the Spanish and for the Chinese findings.

TABLE 2

Configuration comparison tests on frequency-modulated (FM) stimuli based on English intonation contours. Two-way tests for English, Spanish, and Chinese subjects. Results for all contours, falls (HL) only, and rises (LH) only. See Table 1 for explanation

	$r_{VW} = 0$		$A = 1$		$B = 0$		
Test	*df*	r_{VW}	*A*	t_A	*B*	t_B	***CCT Result***
All stimuli							
English versus Spanish (A'-K')	53	.96**	0.96	1.08	0.04	1.06	++
English versus Chinese (A'-K')	53	.95**	0.95	1.13	0.05	1.13	++
Spanish versus Chinese (A'-K')	53	.96**	0.96	1.02	0.04	1.01	++
Falls (Group HL)							
English versus Spanish (A'-G')	19	.96**	0.96	0.61	0.04	0.60	++
English versus Chinese (A'-G')	19	.96**	0.96	0.65	0.04	0.54	++
Spanish versus Chinese (A'-G')	19	.96**	0.96	0.58	0.03	0.57	++
Rises (Group LH)							
English versus Spanish (H'-K')	4	.99**	0.99	0.10	0.00	0.00	++
English versus Chinese (H'-K')	4	.99**	0.99	0.09	0.02	0.00	++
Spanish versus Chinese (H'-K')	4	1.00**	1.00	0.00	0.00	0.00	++

** $p < .01$

TABLE 3

Configuration comparison tests on English intonation contours presented in speech and as frequency-modulated (FM) sine wave stimuli. Separate tests for English, Spanish, and Chinese subjects. Results for all contours, falls (HL) only, and rises (LH) only. See Table 1 for explanation

	$r_{VW} = 0$		$A = 1$		$B = 0$		
Test	*df*	r_{VW}	*A*	t_A	*B*	t_B	***CCT Result***
All stimuli							
English: Speech versus FM	53	.73**	0.73	2.84**	0.26	2.79**	+
Spanish: Speech versus FM	53	.70**	0.70	3.07**	0.30	3.01**	+
Chinese: Speech versus FM	53	.84**	0.84	2.11	0.15	2.07	++
Falls (Group HL)							
English: Speech versus FM	19	.38	–	–	–	–	
Spanish: Speech versus FM	19	–.07	–	–	–	–	
Chinese: Speech versus FM	19	.57**	0.57	2.27	0.43	2.23	++
Rises (Group LH)							
English: Speech versus FM	4	.76	–	–	–	–	
Spanish: Speech versus FM	4	.82*	0.82	0.63	0.18	0.62	+
Chinese: Speech versus FM	4	.95**	0.95	0.31	0.05	0.31	+

** $p < .01$; * $p < .025$; – = test irrelevant

3.3 *Discussion*

For speakers of all three languages, the perception of the FM stimuli duplicated that of the speech stimuli in distinguishing Groups HL and LH. Due to this basic distinction, the correlations between speech and FM judgments across all stimuli were significant for the English, Spanish, and Chinese data.

Only the English and the Spanish FM results differed significantly from the speech results for Group HL. The difference was statistically indistinguishable for the Chinese data. The Chinese listeners seemed to treat the English speech contours more like FM stimuli than did speakers of the other two languages.

The Spanish and Chinese data showed an indistinguishable difference between speech and FM for Group LH. The difference for the English results is probably due to the small number of stimuli. The three FM stimuli H', I', and K' in the LH group formed a tightly bound MDS group. In contrast to the results of Experiment 1, however, the LH stimulus J' was less closely related to the other stimuli in the LH set. This observation applies to all three languages.

Stimulus J' may be separated perceptually in nonspeech because it alone involves only a single pitch movement, a final rise from a very long initial steady-state F0. This particular property of J' seems to be less obvious in its speech version J, which is labeled L* L*H H%. Stimulus J sat one end of the perceptual MDS configuration for Group LH for all listeners, but it remained clearly linked to the other rising contours. Its FM counterpart J', however, was always more weakly linked to the other rising FM contours. The shift in the relative perceptual position of J' as against J may be contingent on the much richer spectral content of speech.

4 General Discussion

Our results show that universal auditory mechanisms that process frequency modulations contribute to the perception of similarities and differences between fundamental frequency contours. Experience with a specific language, however, also contributes to these perceptual processes. We will discuss evidence for these two claims in order.

English, Spanish, and Chinese listeners make a fundamental perceptual distinction between terminal falling (Group HL) and rising (Group LH) F0 contours in speech and nonspeech. Configuration comparison tests using all 11 speech stimuli showed high correlations between the MDS arrays for speakers of the three languages. For the English/Spanish and Spanish/Chinese comparisons, the configurations were indistinguishable in shape. For the perception of the 11 FM analogs of the speech contours, all three cross-language comparisons showed that the configurations were indistinguishable. Configuration comparison tests on the FM stimuli with terminal falls and on the FM stimuli with terminal rises yielded no distinguishable differences between speakers of the three languages. Furthermore, within each language, the correlation between speech and FM configurations across all 11 stimuli was always significant. The two configurations were statistically indistinguishable in the Chinese data. The falling/rising differentiation seems to rest on universal auditory mechanisms that are not language-specific.

Language-specific effects began to emerge in the English/Chinese comparison across all 11 speech stimuli. The speech configurations were similar but not statistically identical. In contrast to the results for the FM stimuli, cross-language differences were clearly evident in the perception of the Group HL speech stimuli. Configuration comparison tests revealed statistical pairwise differences between English, Spanish, and Chinese listeners on the HL speech stimuli. The Chinese subjects placed stimulus G closer to the rising contours than did subjects who spoke the other two languages. This stimulus has a relatively long initial F0 plateau followed by a rise. The differences between the Chinese and the Spanish remained even after elimination of stimulus G. The English/Chinese difference for Group HL, however, disappeared.

All three cross-language comparisons produced perceptual differences for the Group LH rising speech stimuli. The correlations failed to reach significance. This result, however, may well reflect loss of statistical power. Experiments with a larger selection of rising contours might well yield different results.

Data from English and Spanish speakers produced statistically reliable differences between the perception of the speech and the FM stimuli. The differences appeared in the analyses for all stimuli and for the Group HL stimuli. Across all stimuli, the configurations were similar but statistically distinguishable. The configurations proved completely different for the Group HL stimuli in the data for Spanish and for English. The English data also yielded an apparent difference between speech and FM for the four stimuli in Group LH. This result, however, probably reflects low statistical power.

Finally, no reliable differences appeared between any speech and FM configurations for the Chinese results. Speakers of Chinese treated the speech stimuli as more like the FM stimuli than did speakers of English or Spanish. Prosodically, Chinese, a tone language, is different from English and Spanish, which are intonation languages. Consequently, the Chinese listeners should have had the least opportunity to make use of any relevant linguistic-prosodic experience when they judged the stimuli.

In summary, universal auditory mechanisms for processing frequency modulation seem primarily responsible for the basic perceptual distinction between falling and rising contours of intonation in speech. Experience with a given native language then builds on the outputs of those mechanisms in its own particular way. This experience leads to some degree of cross-language specificity in the perception of similarities and differences of intonation contours used in a given language.

Received: October 15, 2002; first revision received: March 19, 2003; second revision received: June 20, 2003; accepted: July 15, 2003

References

ABRAMSON, A. S., & LISKER, L. (1973). Voice timing perception in Spanish initial stops. *Journal of Phonetics*, **1**, 1–8.

ABRAMSON, A. S., & LISKER, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967 (pp. 569–573). Prague: Academia Publishing House of the Czeckoslovack Academy of Sciences.

ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model, *Psychological Science*, **5**, 144–151.

BARTELS, C., & KINGSTON, J. (1994). Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt (Eds.), *Proceedings of the Journal of Semantics Conference on Focus* (pp. 1–10). IBM Working Papers TR-80.94-006.

BEAUGENDRE, F., HOUSE, D., & HERMES, D. J. (2001). Accentuation boundaries in Dutch, French and Swedish. *Speech Communication*, **33**, 305–318.

BECKMAN, M. E. (1986). *Stress and nonstress accent*. Dordrecht: Foris.

BECKMAN, M. E., & AYERS ELAM, G. (1997). Guidelines for ToBI labeling, version 3. Linguistics Department, Ohio State University. *http://ling.ohio-state.edu/~tobi/ame_tobi/*

BEST, C. T., McROBERTS, G., & SITHOLE, N. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English speaker adults and infants. *Journal of Experimental Psychology*, **14**, 345–360.

BLADON, R. A., & LINDBLOM, B. (1981). Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America*, **69**(5), 1414–1422.

BOERSMA, P., & WEENINK, D. (1996). PRAAT: A system for doing phonetics by computer. Report 132 of the Institute of Phonetic Sciences, University of Amsterdam.

BURNHAM, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, **7**(3), 207–240.

CARPENTIER, F., & MOULINES, E. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* **9**, 453–457.

CHEN, A., RIETVELD, T., & GUSSENHOVEN, C. (2001). Language specific effects of pitch range on the perception of universial intonational meaning. *Proceedings of the 9th Eurospeech*, Vol. 2, 1403–1406.

CRUTTENDEN, A. (1997). *Intonation*. Cambridge, U.K.: Cambridge University Press.

CRUZ-FERREIRA, M. (1984). Perception and interpretation of non-native intonation patterns. In M. P. R. van der Broecke & A. Cohen, Eds. *Proceedings of the 10th ICPhS*. Utrecht. Dordrecht: Foris, 565–569.

CUTLER, A., & CHEN, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, **59**(2), 165–179.

DUPOUX, E., PALLIER, C., SEBASTIÁN, N., & MEHLER, J. (1997). A distressing deafness in French? *Journal of Memory and Language*, **36**, 406–421.

EDWARDS, A. L. (1984). *An introduction to linear regression and correlation (2nd ed.)*. New York: Freeman.

EVERITT, B. S., & DUNN, D. (1991). *Applied multivariate data analysis*. London: Arnold.

FLEGE, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

FLEGE, J. E., & HILLENBRAND, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, **79**(2), 508–517.

FLEGE, J. E., MUNRO, M. J., & FOX, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, **95**(6), 3623–3641.

FOX, R. A., & LEHISTE, I. (1987). Discrimination of duration ratios by native English and Estonian Listeners. *Journal of Phonetics*, **15**, 349–363.

FOX, R. A., & LEHISTE, I. (1989). Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners. *Journal of Phonetics*, **17**, 167–174.

GOTTFRIED, T. L., & SUITER, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, **25**, 207–231.

GRABE, E. (1998). *Comparative intonational phonology: English and German*. MPI Series 7, Wageningen: Ponsen en Looien.

GRABE, E., NOLAN, F., & FARRAR, K. J. (1998). IViE—a comparative transcription system for intonational variation in English. *Proceedings of the 5th Conference on Spoken Language Processing*. CD-ROM. (IEEE ICASSP Library Series, New York).

GRABE, E., POST, B., & NOLAN, F. (2001). Modeling intonational variation in English. The IViE system. In S. Puppel & G. Demenko (Eds.), *Proceedings of Prosody 2000* (pp. 51–58). Poznan, Poland: Adam Mickiewitz University.

GUSSENHOVEN, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

GUSSENHOVEN, C. (1990). Tonal association domains and the prosodic hierarchy in English. In S. Ramsaran (Ed.), *Studies in the pronunciation of English* (pp. 27–37). New York: Routledge.

GUSSENHOVEN, C. (2002). Intonation and interpretation: Phonetics and phonology. In B. Bel & I. Marlin (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 47–58). Aix-en-Provence: Laboratoire Parole et Langage.

HART, J. 't, COLLIER, R., & COHEN, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.

HERMES, D. J. (1997). Timing of pitch movements and accentuation of syllables in Dutch. *Journal of the Acoustical Society of America*, **102**, 2390–2402.

HERMAN, R., & McGORY, J. T. (1999). Mapping intonational transcribers' tone similarity space. *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 2331–2334). Berkeley: Congress Organizers.

HIRST, D., & Di CRISTO, A. (1998). *Intonation systems. A survey of 20 languages*. Cambridge: Cambridge University Press.

HOUSE, D., HERMES, D., & BEAUGENDRE, F. (1997). Temporal-alignment categories of accent-lending rises and falls. *Proceedings of Eurospeech, 1997*, Vol. 2, 879–882.

HUANG, T. (2001). Tone perception by speakers of Mandarin Chinese and American English. *Ohio State University Working Papers in Linguistics*, 55.

HUME, E., JOHNSON, K., SEO, M., TSERDANELIS, G., & WINTERS, S. (1999). A cross linguistic study of stop place perception. *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 2069–2072). Berkeley: Congress Organizers.

HUME, E., & JOHNSON, K. (2001). *The role of speech perception in phonology*. New York: Academic Press.

KINGDON, R. (1958). *The groundwork of English intonation*. London: Longman.

KOHLER, K. J. (1987a). Categorical pitch perception. *Proceedings of the XIth International Congress of Phonetic Sciences* (Vol. 5, pp. 331–333). Tallinn: Academy of Sciences of the Estonian Soviet Socialist Republic.

KOHLER, K. J. (1987b). The linguistic functions of F0 peaks. *Proceedings of the Xith International Congress of Phonetic Sciences* (Vol. 3, pp. 149–152). Tallinn: Academy of Sciences of the Estonian Soviet Socialist Republic.

KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, **255**, 606–608.

LADD, D. R. (1996). *Intonational phonology*. Cambridge, U.K.: Cambridge University Press.

LEE, Y.-S., VAKOCH, D., & WURM, L. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, **25**, 527–542.

LEHISTE, I., & FOX, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech*, **35**(4), 419–434.

LISKER, L., & ABRAMSON, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967* (pp. 563–567). Prague: Academia.

MIELKE, J. (2001). A perceptual account of Turkish h-deletion. *Ohio State University Working Papers in Linguistics*, 55.

MIYAKAWA, K., STRANGE, W., VERBRUGGE, R., LIBERMAN, A. M., JENKINS, J. J., & FUJIMORA, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, **18**(5), 331–340.

NASH, R., & MULAC, A. (1980). The intonation of verifiability. In L. R. Waugh & C. H. van Schooneveld (Eds.), *The melody of language* (pp. 219–242). Baltimore: University Park Press.

O'CONNOR, J. D., & ARNOLD, G. F. (1973). *Intonation of colloquial English*. London: Longman.

OHALA, J. (1983). Cross-language use of pitch: An ethological view. *Phonetica,* **40**, 1–18.

PEPERKAMP, S., & DUPOUX, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.

PIERREHUMBERT, J., & STEELE, S. (1989). Categories of tonal alignment in English, *Phonetica*, **46**, 181–196.

PISONI, D. B., LIVELY, S. E., & LOGAN, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121–166). Cambridge, MA: MIT Press.

POLKA, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, **97**(2), 1286–1296.

POLKA, L., & WERKER, J. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421–435.

ROSNER, B. S., & PICKERING, J. B. (1994). *Vowel perception and production*. Oxford: Oxford University Press.

SCHIFFMAN, S., REYNOLDS, M. L., & YOUNG, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.

SILVERMAN, K., BECKMAN, M. E., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J., & HIRSCHBERG, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the Second International Conference on Spoken Language Processing* (Vol. 2, pp. 867–870). Banff, Canada.

SOSA, J. M. (1999). *La entonación del Español. Su estructura fónica, variabilidad y dialectologia.* Madrid: Cátedra.

SPSS INC. (1999). *SPSS Version 10.0.* Chicago: SPSS, Inc.

WERKER, J. F., & TEES, R. C. (1992). The organization and reorganization of human speech perception. In M. Cowan (Ed.), *Annual Review of Neuroscience*, **15**, 377–402.

WILLIAMS, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, **21**, 289–297.

Appendix

ToBI transcriptions for the intonation contours shown in Figure 1 and auditory descriptions using terminology from the British School of intonation analysis.

	Group I: HL			*Group II: LH*	
A	H*L– Prenuclear fall	H*L–L% High nuclear fall	H	L*+H Prenuclear rise	L*H–H% Nuclear rise
B	H*L– Prenuclear fall	!H*L–L% Low nuclear fall	I	L*H– Prenuclear rise	L*H–H% Nuclear rise
C	H*L– Prenuclear fall	^H*L–L% Extra high nuclear fall	J	L* Low level head	L*H–H% Nuclear rise
D	H* Prenuclear fall[1]	LH*L–L% High nuclear fall	K	H+L*H– Prenuclear fall-rise	H+L*H–H% Nuclear fall-rise
E	H* Level head	H*L–L% High nuclear fall			
F	H* Level head	!H*L–L% Low nuclear fall			
G	L*HL– Prenuclear fall rise-fall	L*HL–L% Nuclear rise-fall			

[1] The difference between stimuli C, D, and E in Group I corresponds to Gussenhoven's (1984) notion of "linking." Stimulus C has no linking between two H*L accents. In D, we find partial linking. Stimulus E has total linking plus deletion of the intervening low target. The same applies to stimuli H, I, and J in Group II. Stimulus H has no linking, I has partial linking, and J has total linking plus deletion of the intervening high target.