Perception of English Intonation by English, Spanish and Chinese Listeners

Esther Grabe*, Burton Rosner*, José E. García-Albea† and Xaolin Zhou‡
*University of Oxford, United Kingdom
†Departament de Psicologia, Universitat Rovira i Virgili, Tarragona, Spain
‡Department of Psychology and National Laboratory on Machine Learning, Peking University, Beijing, China

ABSTRACT

We tested the effect of native language on the perception of intonation contours. An English utterance was resynthesized with eleven intonation contours. Groups of English, Iberian Spanish, and Chinese listeners rated pairs of the stimuli for degree of difference. The responses were subjected to multidimensional scaling and a new statistical procedure, the configuration comparison test. The test revealed cross-linguistic similarities and significant differences. All listeners divided the stimuli in two groups, one ending in falling pitch and one ending in rising pitch. Within the falling group, significant cross-linguistic differences occurred in the arrangement of stimuli. In a second experiment, English, Spanish, and Chinese subjects listened to frequency-modulated sinewaves that duplicated the fundamental frequency contours used in the first experiment. Listeners again divided the stimuli into falling and rising groups, but no convincing effect of native language appeared. The speech and FM stimuli seemed more closely related perceptually in the Chinese than in the English or Spanish data.

1 INTRODUCTION

Native language affects the perception of the segmental phonetic structure of speech, the perception of stress and tone and the pragmatic impact of intonation. Native language might also affect the perception of intonation. Cross-language studies on this topic, however, are virtually non-existent. We therefore asked whether English listeners and listeners from other language backgrounds would perceive a given, limited set of English intonation contours in the same way. The previously established cross-language effects on speech perception suggested that listeners from different language backgrounds would perceive different patterns of resemblances between pairs of the English stimuli.

2 BACKGROUND

Native language affects an adult listener's perceptual judgments of the segmental structure of speech [1]. The effect develops during the first year of life [2]. Cross-linguistic differences occur in the perception of stops, fricatives, liquids, and vowels and /h/ [e.g. 3,4,5,6,7].

Lehiste and Fox [8] demonstrated effects of native language on the perception of prominence and stress. Beaugendre, House & Hermes [9] found evidence for a native language effect in the perception of accent location. Dupoux and colleagues [10] showed that French listeners are 'deaf' to stress contrasts that Spanish listeners can identify.

The effect of native language on the perception of tone is not yet solidly established. Gottfried and Suiter [11] and Huang [12] found no major cross-linguistic differences in the perception of tone, although Lee, Vakoch, and Wurm [13] demonstrated such an effect.

A small number of cross-language studies have found differences in the pragmatic effect of intonation [14,15]. We have found no work, however, on the perception of intonation across language groups.

3 METHOD

We carried out two multidimensional scaling (MDS) experiments. MDS represents a set of stimuli as points in a space of the lowest possible dimensionality, while preserving the rank-order relationships between the ratings [16]. 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. Gussenhoven [17], Herman and McGory [18], and Huang [12] have employed MDS in previous work on prosody.

3.1 STIMULI

We used 11 different intonation contours. The contours are well attested in Southern British English and can be taken as relatively uncontroversial [17]. The spoken material, *Melanie Maloney*, was always the same sequence of first name and surname. All contours consisted of one intonation phrase, with a prenuclear accent on *Me-* in *Melanie* and a nuclear accent on *–lo-* in *Maloney*. Phonologically, the contours can 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. This movement

ISBN 1-876346-48-5 © 2003 UAB

was either falling (HL) or rising (LH). Figure 1 illustrates the *f0* patterns of the 11 contours. The shaded areas show the location of the stressed syllables in the contour.

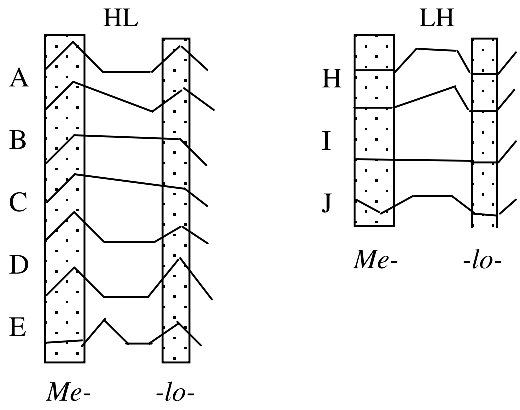

Figure 1. Schematic *f0* patterns for the 11 stimuli. Left and right panels show HL and LH contours, respectively. Shaded areas mark locations of stressed syllables.

A male speaker of Standard British English produced the utterance *Melanie Maloney* with two different contours: a statement, transcribed as H*L H*L %, and a surprised question, transcribed as L*H L*H H%. The experimental stimuli were generated from these two utterances using PSOLA resynthesis implemented in PRAAT 3.8 [19].

3.2 PROCEDURE

Twenty-two native speakers of English, 21 speakers of Spanish and 20 speakers of Mandarin Chinese participated in the experiment. The Spanish and Chinese speakers had rudimentary knowledge of English. Spanish is a stress-accent language like English but has a different inventory of accent shapes [20]. Mandarin is a tone language. Although it has postlexical intonational variation, it does not manifest phrase-long intonational contours constructed from basic pitch movements [21].

The English, the Spanish, and the Chinese listeners were tested in the Phonetics Laboratory at the University of Oxford, the Psycholinguistic Laboratory at the Universitat Rovira i Virgili in Tarragona, and at the Department of Psychology at the University of Peking, respectively. Subject instructions and VDU messages were given in the appropriate language and orthography. Subjects were tested individually in a sound-attenuated room. Stimuli were delivered binaurally over earphones. A program written in C controlled the experiment. The subject first heard each of the 11 stimuli individually. Then a VDU message signaled that self-paced rating trials were about to begin. On each trial one pair of stimuli occurred and was rated from 1 (very similar) to 10 (very different), inclusive. A 200-ms interval intervened between the end of the first stimulus and the start of the second. Each subject rated each of the 110 possible pairs of nonidentical stimuli just once. The pairs occurred in a different random order for each subject. A pair identifier and the rating for each trial were written to a text file. Sessions took around 30 minutes.

The rating data were analyzed with INDSCAL. This multidimensional scaling (MDS) technique is implemented as *PROXSCAL* in the SPSS10.0 statistical package. *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.

4 SPEECH RESULTS

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 raw stress values indicate a good fit of the MDS solutions to the data. Three-dimensional solutions yielded no further improvements.

Figure 2 shows the two-dimensional results. Capital letters (see Figure 1) mark the locations of the individual stimuli in the MDS space. The filled circles represent the weights for the individual subjects, multiplied by 5 for graphical presentation. Points below and above the dashed line in the first quadrant of each plot represent subjects who gave more importance to the horizontal and vertical axes, respectively. The configuration of the weight points in each plot indicates good agreement between listeners.

The English and the Spanish subjects separated the stimuli into the HL and LH groups shown in Figure 1. The Chinese subjects produced a comparable separation, with the possible exception of stimulus G. English and Spanish listeners placed G in the HL group. Chinese listeners moved G into the right half of the MDS space.

4.1 CONFIGURATION COMPARISON TESTS

We assessed cross-linguistic differences by means of a new statistical test, based on regression analysis. The configuration comparison test (CCT) uses all interpoint distances in each of two MDS arrays or subarrays. The distances are invariant under rotations or displacements of the axes. First, one set of distances is normalized in an attempt to match array sizes. Second, a regression line with slope A and intercept B is fitted to the paired distances. If the configurations do not differ, A and B should be 1 and 0, respectively. Third, the correlation r is calculated. If the configurations are similar, r should be well above 0.

Configurations for HL and LH stimuli were analyzed separately. The configurations of the seven falling HL stimuli differed significantly for all language pairings. The value of r never diverged significantly from 0.

The configurations for the rising LH stimuli did not differ between the three sets of subjects, although r was low for the English-Spanish comparison. The small number of degrees of freedom evidently prevented r for this comparison from reaching significance. For the other two LH comparisons, r was large and significant, despite the few available degrees of freedom.

ISBN 1-876346-48-5 © 2003 UAB

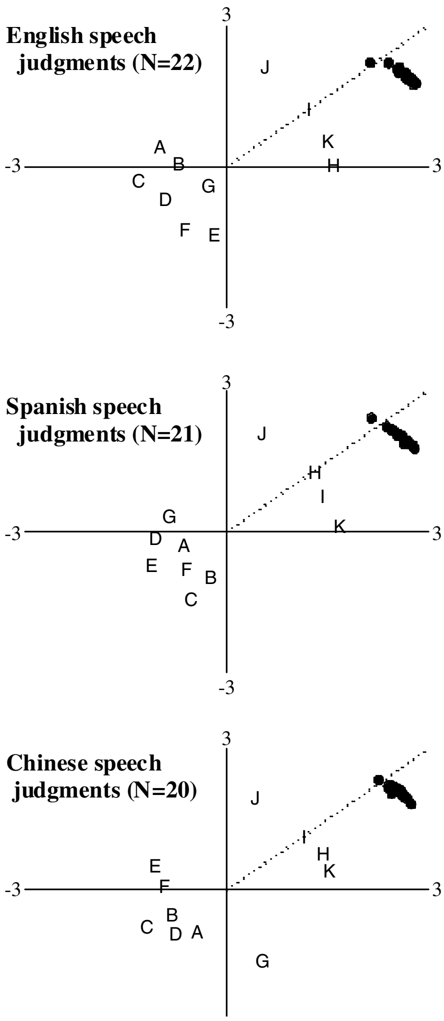

Fig.2 Two-dimensional MDS perceptual maps for ratings of speech stimuli. Letters correspond to contours in Fig. 1.

The perceptual segregation of our stimuli into HL and LH groups by all three sets of listeners suggested that perception of intonation contours might start with the activation of universal auditory mechanisms that process relatively slow frequency modulations. These mechanisms would determine the basic perceptual distinction between the two groups. A listener's native language could modify the outputs of the basic auditory mechanisms, producing different perceptual configurations for stimuli within HL and LH Groups. This hypothesis was investigated in a second experiment.

5 Experiment II

Frequency-modulated (FM) pure tones were made whose contours duplicated those of the speech stimuli, using 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 generated individually and pasted together. Any resulting artifacts were deleted. 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. New English, Spanish, and Chinese listeners took part in the experiment. The testing conditions duplicated those of Experiment 1. The results appear in Figure 3.

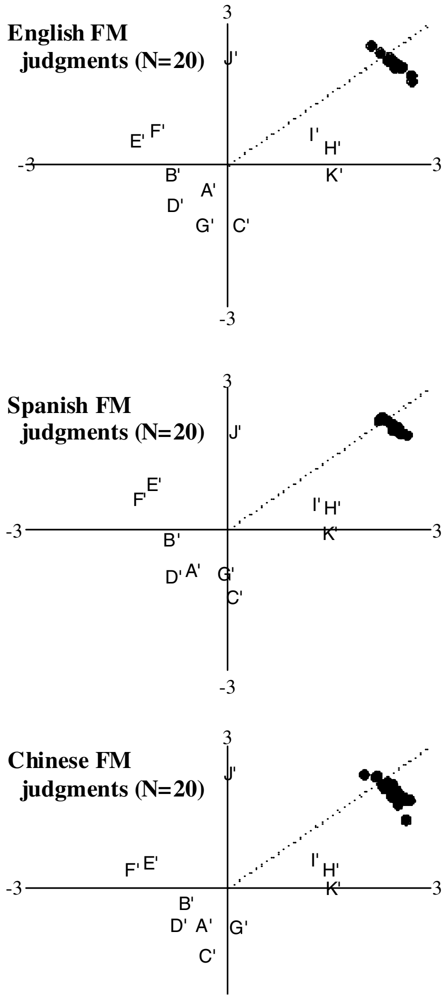

Fig. 3. Two-dimensional MDS map for ratings of FM stimuli by English, Spanish, and Chinese subjects. Primes distinguish FM stimuli from speech stimuli of Exp. 1.

Figure 3 is organized like Figure 2. 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. The plots in Figure 3 confirm our main prediction. All three groups of subjects perceptually separated the stimuli into HL and LH Groups. Furthermore, across languages, the perceptual configurations of stimuli within Groups HL and LH were now much the same. Note that the three configurations for the FM stimuli strikingly resemble the Chinese results for speech. This suggests that the Chinese listeners in Experiment 1 were barely drawing on any linguistic experience.

Configuration comparison tests for the FM stimuli did not demonstrate convincing differences between subject groups. A significant difference emerged for the

ISBN 1-876346-48-5 © 2003 UAB

Spanish-Chinese comparison for stimulus Group HL, and a marginal difference emerged for the English-Chinese comparison. Across all stimuli, the English-Chinese comparison proved significant. In all three instances, however, *B* did not diverge from 0 and *r* was near 1.0. Apparently, compensation for configuration size was inadequate for these three comparisons.

Comparisons of speech and FM judgments within languages revealed significant differences in the HL Groups for English and the Spanish subjects. A marginal difference emerged for Chinese. The correlation between speech and FM judgments was significant for the Chinese but not for the other two sets of results. No difference between speech and FM emerged for the LH Group.

6 CONCLUSION

Our experiments provide evidence for cross-linguistic similarities and differences in the perception of English intonation by English, Spanish and Chinese listeners. Listeners from all three languages divided the stimuli in two groups, one ending in falling pitch and one ending in rising pitch. Within the falling group, we found significant cross-linguistic differences in arrangement of stimuli. In a second experiment, English, Spanish, and Chinese subjects listened to frequency-modulated sinewaves that duplicated the fundamental frequency contours used in the first experiment. Again, listeners divided the stimuli into a falling and rising group. Native language, however, now had no reliable effect. Our findings suggest that universal auditory mechanisms are responsible for processing slower frequency modulations. The perception of variation within a given class of slower modulations, however, may be language-specific.

ACKNOWLEDGMENTS

This work was supported by grant R000237145 from the Economic and Social Research Council, UK, 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 Bob Ladd for helpful suggestions and the late Nancy C. Waugh for numerous useful comments on earlier drafts of this paper.

REFERENCES

[1] L. Polka, “Linguistic influences in adult perception of non-native vowel contrasts”, Journal of the Acoustical Society of America, 97(2), 1286-96.

[2] P.K. Kuhl, K.S. Williams, F. Lacerda, K. N. Stevens and B. Lindblom, “Linguistic experience alters phonetic perception in infants by 6 months of age”, Science, 255, 606-8,1992.

[3] L. Lisker and A. Abramson, “The voicing dimension: Some experiments in comparative phonetics”, Proceedings of the Sixth International Congress of Phonetic Sciences, pp. 563-7, 1970.

[4] J. Flege and J. Hillenbrand, “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), 508-17, 1986.

[5] K. Miyakawa, W. Strange, R. Verbrugge, A. Liberman, J.J. Jenkins and O. Fujimora, “An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English”, Perception & Psychophysics, 18(5), 331-40, 1975.

[6] J.E. Flege, M.J. Munro and R.A. Fox, “Auditory and categorical effects on cross-language vowel perception”, Journal of the Acoustical Society of America, 95(6), 3623- 41, 1994.

[7] J. Mielke, “A perceptual account of Turkish h-deletion”, Ohio State University Working Papers in Linguistics, vol 55, 2001.

[8] Lehiste and R. A. Fox, “Perception of prominence by Estonian and English listeners”, Language and Speech, 35, vol. 4, 419-34, 1992.

[9] F. Beaugendre, D. House and D. J. Hermes, “Accentuation boundaries in Dutch, French and Swedish” Speech Communication, vol 33, pp. 305-318, 2001.

[10] E. Dupoux, C. Pallier, N. Sebastian and J. Mehler, “ A distressing deafness in French?”, Journal of Memory and Language, vol. 36, pp. 406-21, 1997.

[11] T.L. Gottfried and T.L. Suiter, “Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones”, Journal of Phonetics 25, 207-31, 1997.

[12] T. Huang, “Tone perception by speakers of Mandarin Chinese and American English”, Ohio State University Working Papers in Linguistics, 55, 2001.

[13] Y.-S. Lee, D. Vakoch L. Wurm, “Tone perception in Cantonese and Mandarin: A cross-linguistic comparison”, Journal of Psycholinguistic Research, 25, 527- 42, 1996.

[14] M. Cruz-Ferreira, “Perception and interpretation of non-native intonation patterns”, In M. P. R. van der Broecke and A. Cohen, Eds. Proceedings of the 10th ICPhS, pp. 565-9, 1984.

[15] Chen, T. Rietveld and C. Gussenhoven, “Language specific effects of pitch range on the perception of universial intonational meaning”, Proceedings of the 9th Eurospeech, vol. 2, pp. 1403-6, 2001.

[16] S. Schiffman, M.L. Reynolds and F.W. Young, Introduction to multidimensional scaling, Academic Press, 1981.

[17] C. Gussenhoven, On the grammar and semantics of sentence accents, Foris, 1984.

[18] R. Herman and Tevis McGory, J. “Mapping intonational transcribers’ tone similarity space”, Proceedings of the XIVth International Congress of Phonetic Sciences, pp. 2331-4, 1999.

[19] P. Boersma and D. Weenink, “PRAAT: A system for doing phonetics by computer”, Report 132 of the Institute of Phonetic Sciences, University of Amsterdam, 1996.

[20] J.M. Sosa, La entonación del Español. Su estructura fónica, variabilidad y dialectologia. Cátedra, 1999.

[21] D.R. Ladd, Intonational phonology, CUP, 1996.