Applied Psycholinguistics **21** (2000), 383–393
Printed in the United States of America

Same talker, different language

VERNA STOCKMAL, DANNY R. MOATES, and ZINNY S. BOND
Ohio University

ADDRESS FOR CORRESPONDENCE
Zinny Bond, Department of Linguistics, Ohio University, Gordy Hall, Athens, OH 45701.
Email: bond@ohio.edu

ABSTRACT
When discriminating between unknown foreign languages, infants, young children, and adult listeners are able to make same-language/different-language discrimination judgments at better than chance levels. In these studies (Lorch & Meara, 1989; Mehler et al., 1988; Stockmal, 1995), foreign language samples have often been provided by different talkers, confounding voice characteristics and language characteristics. In Experiments 1 and 2, using the same talkers for different pairs of languages, we found that listeners were able to discriminate between languages they did not know, even when spoken by the same talker. That is, listeners were able to separate talker from language characteristics. Experiment 3 used multidimensional scaling to explore the bases of listener judgments. Listeners were attentive to prosodic properties and influenced by their familiarity with the test languages.

Human listeners are able to make reliable language identification judgments from very short speech samples, apparently by employing multiple sources of information (Bond & Fokes, 1991; Fucci, McColl, Bond, & Stockmal, 1997; Lorch & Meara, 1989, 1995; Stockmal, Muljani, & Bond, 1994, 1996). Attempts to develop algorithms for automatic language identification and discrimination judgments have led researchers to examine human performance in order to obtain benchmarks for evaluating machine performance (Muthusamy, Bernard, & Cole, 1994; Muthusamy, Jain, & Cole, 1994). Presumably the primary source of information listeners employ is the "acoustic signature": that is, the unique acoustic–phonetic pattern as well as structural properties of each language (Hombert & Maddieson, 1998; Muthusamy, Bernard, & Cole, 1994).[1]

Listeners may also be influenced by extralinguistic characteristics such as talker voice, gender, and speech style and rate. Talker voice and language characteristics have been shown to interact in the perception of native listeners. For example, Jerger et al. (1993) asked children and adults to attend selectively to either a voice dimension or a linguistic dimension in a word recognition task. They found that irrelevant variation in the voice dimension disrupted performance more than irrelevant variation in the linguistic dimension for all age groups. Palmeri, Goldinger, and Pisoni (1993) reported that words repeated in a familiar voice were classified as "new" or "old" more accurately than words repeated by novel talkers. When listeners were asked to focus on talker voice, same gender word repetitions were more difficult for subjects to identify cor-

© 2000 Cambridge University Press 0142-7164/00 $9.50

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

rectly than different gender repetitions of the same words. Apparently, encoded voice information contains categorization by gender, among other properties. Consistent with these findings, Mullenix, Johnson, Topcu-Durgun, and Farnsworth (1995) argued that phonetic representations are categorical and therefore less detailed in memory than voice representations (e.g., gender), which contain speaker-specific auditory information. Pisoni (1992) suggested that both phonetic information and talker-specific attributes are encoded as an integral part of the perceptual record in long-term memory (p. 144).

Talker voice may be more salient in responding to an unknown foreign language than in responding to a known language. Listeners to unknown languages lack lexical and grammatical information; they have available only phonetic representations and talker-specific information. Moreover, Esling and Wong (1983) suggested that talker voice quality or "voice settings" – a term used to describe the auditory impression made by a certain mechanical setting of the speech organs over a stretch of speech – may function to characterize not only a speaker but also a language or a speaking style associated with a particular geographically or socially defined language variety.

In fact, there is considerable evidence that talker voice characteristics interact with language identification. In a number of studies (Lorch & Meara, 1995; Muthusamy, Jain, & Cole, 1994; Stockmal et al., 1996), listeners were sensitive to talker voice quality and sometimes reported difficulty separating talker voice from language characteristics. Lorch and Meara (1995) reported that their subjects were aware of both talkers and specific language stimuli. Muthusamy, Jain, and Cole (1994) found that some listeners attempted to identify languages by comparing the voice quality of a speech sample with the voice quality of a friend whose native language the listeners knew. Stockmal (1995) found that the same sentence produced in a different voice led to a decrease in the accuracy of same-language judgments. In a multidimensional scaling analysis, Stockmal et al. (1996) interpreted one dimension as talker-specific, representing voice quality and speaking rate.

In all these studies, samples of the different languages have been provided by different talkers. That is, language characteristics and talker characteristics have been confounded. In order to investigate whether listeners can separate talker voice from language characteristics, we employed talkers fluent in two languages. In Experiments 1 and 2, we examined the relationship between talker variability and language identity. In Experiment 3, we investigated the perceptual dimensions on which listeners base their judgments by employing multidimensional scaling.

EXPERIMENT 1

The purpose of the first experiment was to determine if listeners could discriminate between foreign languages when produced by bilingual talkers.

Method

Materials. Four male and four female talkers with native-like proficiency in two languages recorded passages in each language. The passages were taken

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

Table 1. *Language pairs used in the study*

Male talkers	Female talkers
Arabic–French (Morocco)	Korean–Japanese (Asia)
Hebrew–German (Israel)	Ombawa–French (Cameroon)
Akan–Swahili (Africa)	Latvian–Russian (Europe)
Latvian–Russian (Europe)	Ilocano–Tagalog (Philippines)

Note: In each pair the talker's native language is given first.

from newspapers or popular magazines. The talkers read the passages at a self-selected rate. The recordings were made in a language laboratory recording studio. Typically, one language of the talkers was spoken at home and the other was the language the talker used outside the home. The languages represented Africa, Asia, and Europe (see Table 1). The languages differed in segment inventories and suprasegmental properties. On the continuum from syllable to stress rhythm, Arabic, German, and Russian are considered to employ stress rhythm, whereas Akan, French, Ilocano, Korean, Latvian, Ombawa (or Mbawa), Swahili, and Tagalog employ syllable rhythm; Japanese may employ both syllable and mora rhythm. Akan, Ombawa, and Japanese show tone or pitch accent variation over lexical items.

Two test recordings were created using the Kay Elemetrics Computerized Speech Lab. One test recording contained language samples produced by male talkers, and the other, by female talkers. Each recording contained fluent 5-second phrases or sentences excerpted from the recorded passages. Each excerpt appeared only once in the test recordings.

The excerpts were paired with either the same language or the other language spoken by that talker in four different combinations for each language pair. For example, the Japanese–Korean test items consisted of Japanese–Korean, Korean–Japanese, Japanese–Japanese, and Korean–Korean pairings. Excerpts within test items were separated by a short beep. No excerpt was paired with itself. Each pairing of languages occurred twice for a total of 32 test items.

Listeners. A total of 62 undergraduates enrolled in classes in linguistics, psychology, or hearing and speech sciences participated in the experiment for a small amount of course credit. All were native speakers of American English and reported speech, language, and hearing abilities within normal limits. No listener had fluency in, or extensive exposure to, any of the test languages.

Procedure. Listeners were tested in small groups. They were instructed to listen to each test item, decide whether the excerpted samples were spoken in the same language or in different languages, and respond by indicating "same" or "different" on a prepared answer sheet. They were not told what the languages were. Listeners heard only one recording, produced by either male or female talkers.

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

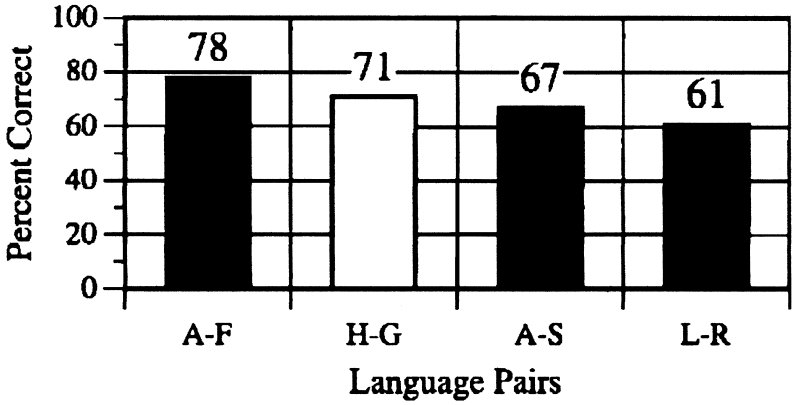


Figure 1. Percent correct judgments for the four language pairs produced by the male talkers.

Results

A nonparametric measure of sensitivity, A', was calculated for each listener. A' is a descriptive statistic used to evaluate listener sensitivity. In same/different judgments, listeners may make an error by failing to detect differences between stimuli (responding "same" when they should say "different") or by failing to detect similarities (responding "different" when they should say "same"). A' reflects both types of error. It is analogous to the d' used to measure sensitivity in signal detection. The value of A' ranges from 0 to 1, with 0.5 representing chance performance (Grier, 1971; McNichol, 1972; Noreen, 1981). A' is calculated on the basis of "hits" and "false alarms." In this experiment, correct "different" judgments were defined as "hits," and same-language pairs judged as different were considered "false alarms."

The mean A' value across all listeners for male talkers was 0.77 ($SD = .09$). The corresponding mean value for all female talkers was 0.71 ($SD = .11$). These values were significantly above chance performance both for male talkers, $t(27) = 15.88$, $p < .001$, and for female talkers, $t(33) = 37.57$, $p < .001$.

Discussion

The acoustic signature of the languages provided sufficient information for the listeners to discriminate between languages they did not know, even when the same talker provided the spoken samples. This finding indicates that listeners were able to separate voice quality characteristics from language-specific, acoustic–phonetic characteristics. In a similar task, Stockmal (1995) reported 79% correct discrimination of different languages produced by different female talkers. The mean percent correct in the present study was somewhat lower: 69% for male talkers and 66% for female talkers. The mean percent correct in this study was based on both "hits" and "correct rejections."[2]

Not all language pairs were equally discriminable. As can be seen in Figures 1 and 2, all different-language pairs for male talkers were discriminable,

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

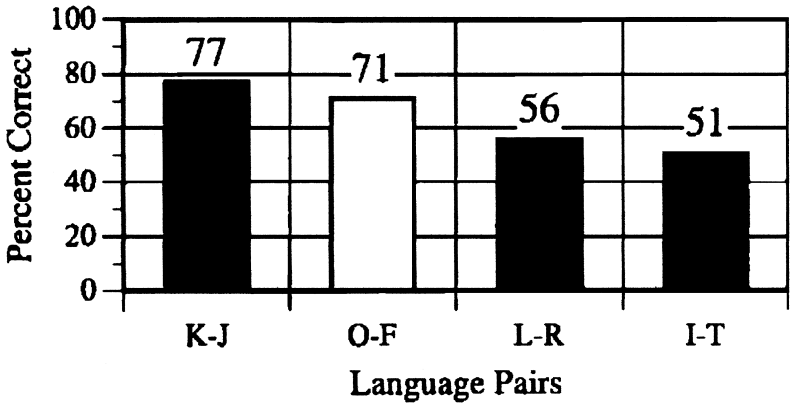


Figure 2. Percent correct judgments for the four language pairs produced by the female talkers.

whereas only the Japanese–Korean and French–Ombawa pairs for female talkers were clearly discriminable. The Latvian–Russian pairs produced by the male talker were the most difficult language pairs on the male talker test. The Latvian–Russian pairs produced by the female talker were discriminated nearly at chance, and the Ilocano–Tagalog pairs received essentially chance responses. One possible explanation for the confusability of Latvian and Russian is that both languages share the phonotactic properties of employing complex syllable onsets. In turn, Ilocano and Tagalog are structurally similar in terms of segment inventories, prosodic patterns, and voice quality/setting. The listeners could not employ morphological and lexical information, and the acoustic–phonetic properties of the languages provided insufficient differences for their discrimination.

The most easily discriminated language pairs in the male talker set included at least one somewhat familiar language – French, Hebrew, or German. Although French also appeared in one of the two discriminable pairs of languages produced by the female talkers, the two most discriminable language pairs each contained one language with either tone or pitch accent. Pitch variation appears to be one of the most salient characteristics available to listeners in a language discrimination task (Bond, Stockmal, & Muljani, 1998; Lorch & Meara, 1989).

To provide listeners with voice differences that would increase the variability of talkers, a second experiment was conducted in which stimuli were presented in both male and female voices.

EXPERIMENT 2

Method

Materials. A recording was created from the speech samples of the two male and two female talkers receiving the highest A' scores in Experiment 1. The

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

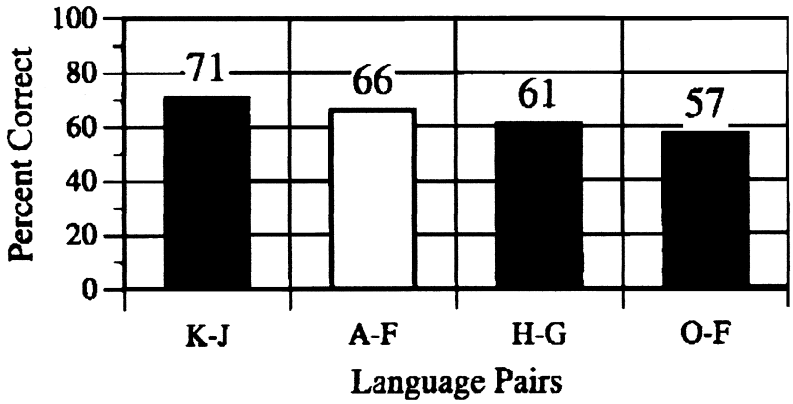


Figure 3. Percent correct judgments for the four language pairs produced by male and female talkers.

language pairs employed for the male talkers were Arabic–French and Hebrew–German; the language pairs for the female talkers were Korean–Japanese and Ombawa–French. In other respects, the recording was identical to those used in Experiment 1.

Listeners. A total of 69 students participated in the second experiment. They were selected from the same population as the listeners in the first experiment. None had participated in the first experiment.

Procedure. The procedure was the same as in the first experiment.

Results

As in Experiment 1, A' was calculated for each listener. The mean value of A' for all talkers was .71 ($SD = .11$). This value was significantly above chance, $t(68) = 16.15$, $p < .001$, indicating that the listeners were able to discriminate spoken samples of foreign languages when produced by both male and female talkers in a more demanding listening condition.

Percent correct was calculated for each different-language pair and is given in Figure 3. All pairs were discriminable, although listener performance for the language pairs was less accurate than in the first experiment. The difference in correct judgments between the two experiments was approximately 10%.

Discussion

Experiment 2 showed that the findings of Experiment 1 were reliable. Listeners were able to detect distinctive acoustic patterns for the unknown languages and to abstract away from the category "talker" to the category "language" on the basis of the phonetic information present in the speech signal. Overall, listener

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

scores were lower than in Experiment 1, suggesting that listening to male and female talkers was more difficult than listening to a same-gender set. These findings lend support to the suggestion that increased variability on one dimension affects decisions on the other dimension (Pisoni, 1992).

As in Experiment 1, not all language pairs were equally discriminable. The Korean–Japanese pairs, consisting of a language (Japanese) which uses pitch accent paired with a language which does not employ much pitch variation, were the easiest to discriminate. The least discriminable languages employed similar rhythmic organization. Taken together with the results of Experiment 1, this finding suggests that the information contained within the suprasegmental properties of these languages serves as one basis of discrimination.

The test language samples employed in Experiments 1 and 2 were analyzed in a same/different format that does not reveal the perceptual dimensions listeners used for judgments. Experiment 3 used multidimensional scaling analysis (MDS) to investigate the perceptual bases of same/different judgments. The aim of MDS is to find structure in a data set. MDS procedures represent objects judged experimentally similar to one another as points close to each other in a resultant spatial map. MDS procedures do not require a priori knowledge of the attributes of the objects in the data set (Schiffman, Reynolds, & Young, 1981, p. 3).

EXPERIMENT 3

Method

Materials. Two different randomizations of the test recording used in Experiment 2 were used in Experiment 3.

Listeners. A total of 52 students participated in this experiment. They were selected from the same population as the listeners in the first two experiments. None had participated in the first two experiments.

Procedure. Listeners were instructed to listen to the language pairs and to rate the similarity of the languages in each pair using a 7-point scale (Schiffman et al., 1981; Young & Harris, 1997). Listeners assigned a rating of 1 to an item if they considered the languages to be very different and a rating of 7 if they considered the languages to be very similar or identical. Several practice trials using Latvian–English language pairs were given until the listeners were comfortable with the procedure. Each listener then rated the 64 experimental language pairs.

Results

The mean rating for same-language pairs was 5.19 ($SD = .51$), which was significantly different from the different-language pair value of 3.45 ($SD = .57$), $t(51) = 20.43$, $p < .001$.

A MDS was performed on the similarity ratings for both the same- and differ-

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

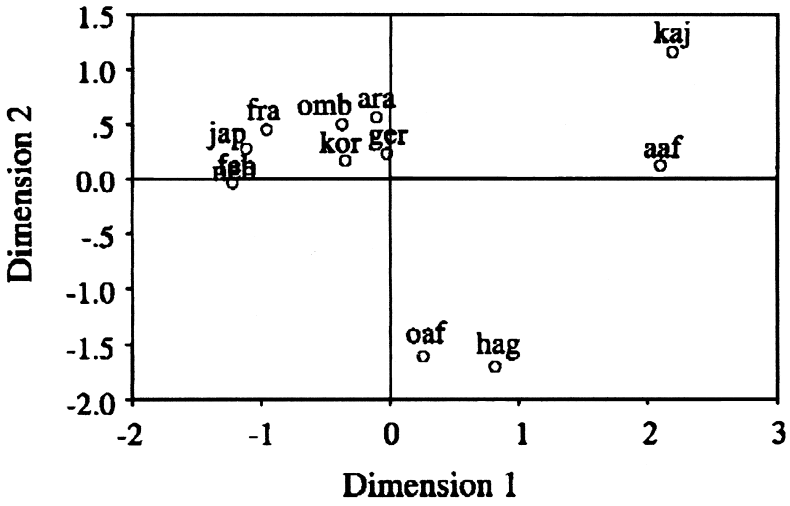


Figure 4. Two-dimensional solution for the multidimensional scaling analysis of judgments on the same- and different-language pairs. The same-language pairs, located in the top left quadrant, are identified by a three-letter code: ara = Arabic; fra = French male; frb = French female; ger = German; heb = Hebrew; jap = Japanese; kor = Korean; omb = Ombawa. The different-language pairs are located in the right half of the map: aaf = Arabic and French; hag = Hebrew and German; kaj = Korean and Japanese; oaf = Ombawa and French.

ent-language pairs. The best solution for the dissimilarity matrix for the eight languages was computed by ALSCAL as a two-dimensional Euclidian space, RSQ = .997. The two-dimensional solution, given in Figure 4, was chosen on the basis of variance accounted for and interpretability.

Discussion

Neither talker gender nor voice emerged as a dimension in the MDS analysis. Dimension 1, the horizontal axis, corresponds to listeners' perception of language similarity within each stimulus pair. Same-language pairs fall to the left of the midpoint, and different-language pairs fall to the right. Dimension 2, the vertical axis, corresponds to the ease with which language pairs could be discriminated. Within the set of same-language pairs, there is essentially no variation on this dimension. Within the set of different-language pairs, increases along the vertical axis correspond to increases in the ease with which the language pairs could be discriminated. The easiest to discriminate were the language showing pitch excursion (Japanese) versus the relatively flat pitch pattern language (Korean). Somewhat more difficult to discriminate were the stress versus syllable-timed language pairs (Arabic and French). The most difficult language pairs contained two syllable-timed languages (French and Ombawa). Both of these languages also employ vowel nasalization.

Within the same-language set, a pattern of subsets defined by prosody and

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

familiarity emerged. Relevance of prosodic pattern to perceptual organization is indicated by the fact that the two stress-timed languages within the set clustered together. The proximity of Hebrew to French may result from the phrase-final primary stress placement characteristic of both languages. French, Hebrew, and also Japanese – all somewhat familiar to our listeners – form a relatively tight cluster, whereas the cluster of less familiar languages is loosely configured. Though some of these groupings are related to traditional notions of rhythm, the segmental inventories and phonotactics are undoubtedly related to judgments as well (Dauer, 1983, 1987; Hombert & Maddieson, 1998).

GENERAL DISCUSSION

Experiments 1 and 2 showed that listeners were able to discriminate unknown languages at better than chance rates: that is, to distinguish foreign languages on the basis of phonetic information present in the speech signal. The results of the MDS analysis (Experiment 3) suggest that listeners were employing prosodic information within the context of language familiarity to make the discriminations. Ideally, all languages employed in the experiments would have been completely unfamiliar to the listeners. Unfortunately, for many of the talkers, one of their languages was a language also offered at American universities and so tended to be somewhat familiar to American college students. Although language familiarity affected listener responses, language pairs consisting of unfamiliar languages such as Akan and Swahili were also discriminable, suggesting that listeners could make use of phonological information from completely unfamiliar languages. For all languages, sensitivity to prosodic pattern is evident in judgments about both same- and different-language pairs.

Early automatic language identification studies have had only marginal success using prosodic information in language identification (Muthusamy, Bernard, & Cole, 1994). However, prosody defined by linguistic parameters and employed in algorithms used in these studies may not have included enough of the complex information encoded as prosody by human listeners. For human listeners, prosody may be considered a "complex of voice quality, intonation, rhythm, stress and phonological fluency" (Pennington, 1996, p. 163). According to Esling (1987), "voice setting acts as the background or field, stress and intonation are overlaid as prosodic waves on that field, and phonemic segments appear as particles in rapidly changing succession, bearing a relationship to both the medium-term (prosodic) and long-term background (setting) elements" (p. 45). In this study, even though bilingual talkers were employed, their native-like ability in both languages sometimes produced noticeably different voice settings for the different languages. For example, the Korean talker tended to speak Japanese at a higher pitch than her native language. For a number of the talkers, their two languages represented more than simply alternate ways of communicating a message; specific languages were associated with different aspects of their life experiences. However, basic voice characteristics of the talkers, such as timbre and gender, remained constant in their two languages. Clearly, listeners were able to abstract from talker to language characteristics to make discrimination judgments.

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

ACKNOWLEDGMENTS

Preliminary versions of portions of this article were presented at the Midwestern Psychological Association, Chicago, 1998, and at the International Congress on Spoken Language Processing, Sydney, 1998. We acknowledge with gratitude the talkers who made this project possible. We also appreciate the comments of Scott Jarvis on an earlier version of this article.

NOTES

1. The distinction between acoustic–phonetic properties and linguistic properties or properties based on meaning is pervasive in the automatic speech recognition literature. Essentially acoustic–phonetic properties can be measured from the physical signal.
2. Because this experiment focused on discrimination, same-language pairs judged as "same" were defined as "correct rejections."

REFERENCES

Bond, Z. S., & Fokes, J. (1991). Identifying foreign languages. *Proceedings XII Congress of Phonetic Sciences* (pp. 198–201).

Bond, Z. S., Stockmal, V., & Muljani, D. (1998). Learning to identify a foreign language. *Language Sciences, 20*, 353–367.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51–62.

(1987). Phonetic and phonological components of language rhythm. *Proceedings XI Congress of Phonetic Sciences* (pp. 447–450).

Esling, J. H. (1987). Methodology for voice setting awareness in language classes. *Revue de Phonetique Appliquee, 85,* 449–473.

Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17*, 89–95.

Fucci, J. H., McColl, D., Bond, Z. S., & Stockmal, V. (1997). Magnitude-estimation scaling of complex auditory stimuli: A comparison of native and non-native speakers of English. *Perceptual and Motor Skills, 85*, 1468–1470.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424–429.

Hombert, J. M., & Maddieson, I. (1998). A linguistic approach to automatic language recognition. *UCLA Working Papers in Phonetics, 96*, 119–125.

Jerger, S., Pirozzolo, F., Jerger, J., Elizondo, R., Desai, S., Wright, E., & Reynosa, R. (1993). Developmental trends in the interaction between auditory and linguistic processing. *Perception and Psychophysics, 54*, 310–320.

Lorch, M., & Meara, P. (1989). How people listen to languages they don't know. *Language Sciences, 11*, 343–353.

(1995). Can people discriminate languages they don't know? *Language Sciences, 17*, 65–71.

McNichol, D. (1972). *A primer of signal detection theory*. London: George Allen & Unwin.

Mehler, J., Jusczyk, P., Lampertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.

Mullenix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 84*, 3080–3095.

Muthusamy, Y. K., Bernard, E., & Cole, R. (1994). Automatic language identification: A review tutorial. *IEEE Magazine*, 1–16.

Muthusamy, Y. K., Jain, N., & Cole, R. (1994). Perceptual benchmarks for automatic language identification, *Proceedings IEEE International Congress on Acoustics, Speech, and Signal Processing* (pp. 1333–1336).

Noreen, D. (1981). Optimal decision rules for common psychophysical paradigms. In S. Grossberg, (Ed.), *SIAM Proceedings V* (pp. 238–281). Hillsdale, NJ: Erlbaum.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press

and recognition memory for spoken words, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309–328.

Pennington, M. C. (1996). *Phonology in English language teaching*. London: Longman.

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 143–151). Tokyo: Ohmsha.

Schiffman, S., Reynolds, M. L., & Young, F. (1981). *Introduction to multidimensional scaling*. Orlando, FL: Academic.

Stockmal, V. (1995). *Discrimination of unknown foreign languages in spoken utterances: A developmental study*. Unpublished master's thesis, Ohio University, Athens.

Stockmal, V., Muljani, D., & Bond, Z. S. (1994). Can children identify samples of foreign languages as same or different? *Language Sciences, 16*, 237–252.

(1996). Perceptual features of unknown foreign languages as revealed by multidimensional scaling. *Proceedings International Congress on Spoken Language Processing* (pp. 1748–1751).

Youn F. W., & Harris, D. F. (1997). *Multidimensional Scaling in SPSS Professional Statistics 7.5*. Chicago: SPSS Inc.

https://doi.org/10.1017/S0142716400003052 Published online by Cambridge University Press