

ON TALKER VOICE AND LANGUAGE IDENTIFICATION

Verna Stockmal, Danny Moates & Z. S. Bond

Ohio University, Athens, Ohio, USA

vstockmal@gmail.com; moates@ohio.edu; bond@ohio.edu

ABSTRACT

Listener similarity judgments of languages seem to be influenced by regional speech characteristics and talker voice quality, and listener responses to voice quality are influenced by language. This study attempted to assess the relationships between judgments about voice quality and judgments about language. In the first experiment, using an ABX format listeners matched spoken samples of unknown languages when produced by male and female talkers. Overall, listeners performed at above chance level. In the second experiment, listeners rated the similarity of the same talkers. Listeners found talkers most similar when they were paired with themselves. They judged talkers speaking the same language as more similar than talkers speaking a different language, even across gender.

Keywords: voice quality, language identification, Arabic, Latvian

1. INTRODUCTION

Listener ability to discriminate or identify spoken samples of foreign languages which they neither speak nor understand has been demonstrated in several studies [1, 5, 6]. When hearing an unknown language, listeners must rely entirely on phonetic information for judgments about it. According to Pisoni and colleagues [7, 10, 11], the perceptual processes used to remember phonetic information and the processes used to remember talker voice information do not function independently; rather, detailed information about talker voice and linguistic properties is encoded together in memory. Talker voice information can also be expected to be encoded with phonetic information of unknown languages.

In some studies of language discrimination, researchers have suggested that talker voice affects listener judgments [1, 6]. Muthusamy & Cole [9] speculated that listeners sometimes even try to identify an unknown language on the basis of voice characteristics of persons known to speak it. Nevertheless, listeners are able to discriminate

between spoken samples of unfamiliar languages produced by proficient bilingual talkers [13, 14].

The purpose of this study is to investigate the relationship between talker voice characteristics and the phonetic characteristics of languages not known to listeners. The first experiment examined listener ability to match languages when samples were produced by male and female talkers; the second experiment investigated the influence of language on voice in similarity judgments. The languages employed were Arabic and Latvian.

2. EXPERIMENT 1

Because listeners may use different perceptual strategies when classifying male and female voices [8, 12], it is possible that talker gender interacts with listener ability to identify spoken samples of unknown languages. The objective of the first experiment was to determine to what extent naïve listeners can identify a language as the same when produced by male and female talkers.

2.1. Method

2.1.1. Participants

Eighteen American college students with self reported normal speech and hearing served as listeners. None had any experience with the languages used in the test.

2.1.2. Materials

Three female and three male talkers of Latvian and three female and three male talkers of Arabic recorded a self-selected short prose passage at a normal reading rate. All the Latvian talkers were residents of Riga. The Arabic talkers came from several different areas, including Saudi Arabia, Morocco, Egypt, Syria, and Palestine. Phrases were arranged in ABX format, paired across both language and gender. Both languages and both genders appeared in A, B and X positions but neither a phrase nor talker was repeated within a triad.

2.1.3. Procedure

Participants heard the test recording in a quiet classroom. They were instructed to decide whether the language in the X position matched the language in either the A or B position.

2.2. Results

Overall, listeners could do the task. They matched both languages equally well at 62% correct, significantly better than chance expectation [$t = 5.48$, $p < .001$]. Listeners matched male talkers with male talkers more accurately than female talkers with female talkers, 73% vs. 52% correct [$F(3, 68) = 11.38$, $p < .0001$]. They matched languages somewhat less accurately when talker gender differed, 59% correct. The interaction between language and talker gender was not significant. Correct matches of X with samples in either the A or B position also did not differ significantly.

There was considerable variability in the number of correct responses to specific test items, from 17% to 94% correct. Arabic as produced by females was somewhat easier to match than Arabic produced by males while Latvian produced by males was somewhat easier than Latvian produced by females.

2.3. Discussion

Listeners could abstract language from talker voice and match language samples correctly. Nevertheless, talker voice influenced responses, at least in the sense that female voices were more difficult to match to targets provided by either male or female voices.

The two languages did not seem to differ in identifiability even though the Arabic talkers represented different varieties of the language whereas all the Latvian talkers were residents of the same city, Riga. The Latvian talkers ranged in age more than the Arabic talkers did but these differences also did not systematically affect listener judgments.

3. EXPERIMENT 2

The second experiment examined the influence of language when listeners were attending to voice. Because listeners have difficulty identifying voices in a language they do not understand [4] scaling was selected for cross-linguistic comparisons.

3.1. Method

3.1.1. Participants

One hundred thirteen American English monolingual college students served as listeners. All had self reported normal speech and hearing. No participant had ever lived in an environment where different languages were spoken or could speak any foreign language. None had participated in the previous experiment.

3.1.2. Materials

The same six native speakers of Latvian and Arabic provided the language samples. These samples were assembled into four test tapes. Tape A paired only female voices. Each female talker was paired with herself, with each of the other two talkers of her native language, and with each talker of the other language for a total of 42 test items. Tape B paired only male voices in the same way. Tape C paired the females from both languages with all the male speakers of Arabic for a total of 36 test items. Tape D paired all females with all male speakers of Latvian.

3.1.3. Procedure

Listeners were divided into four groups. Each group responded to only one test recording. The listeners were asked to rate the similarity of the two voices in each test item on a scale from one to ten. One was the rating if listeners thought the two voices were the same; ten was the rating for the most dissimilar voices. No specific instructions were given on how listeners were to judge similarity. A questionnaire was administered after the listening test in which listeners provided information about their background in language study and the bases for their similarity judgments.

3.2. Results

3.2.1. Same talkers

Overall, listeners were able to recognize the same talker in unknown languages. In the subset of ratings in which listeners heard two samples of speech produced by the same talker, the mean dissimilarity score was less than 3. There was no significant difference in ratings for male and female talkers.

3.2.2. Same gender talkers

Similarity judgments for male talkers were submitted to a repeated measures analysis of variance. Language and talker were the two independent variables. For male talkers, there was a significant interaction between talker and language [$F(5, 1074) = 35.51, p < .01$]. The main effects for talker and for language were also significant [$F(5, 1074) = 5.03, p < .01$; $F(1, 1074) = 553.16, p < .01$]. Talkers speaking the same language were judged to be more similar to each other than talkers speaking a different language, suggesting that language influenced similarity judgments.

Similarity judgments for female talkers were also submitted to a repeated measures analysis of variance. As in the male analysis, language and talker were the two independent variables. For female talkers, there was a significant interaction between talker and language [$F(5, 1074) = 45.42, p < .01$]. The main effects for talker and for language were also significant [$F(5, 1074) = 64.97, p < .01$; $F(1, 1074) = 24.224, p < .01$]. For all but the Latvian females, talkers were rated as more similar when they spoke the same language, regardless of other voice characteristics.

3.2.3. Cross-gender talkers

Similarity ratings for the two groups of female talkers, paired with the Arabic males were compared in a *t* test for independent samples. Results showed that Arabic females were judged significantly more similar to Arabic males than were Latvian females [$t = 3.98, p < .01, 2$ -tailed]. When Latvian males were paired with both groups of females, the Latvian females were judged significantly more similar to the males than the Arabic females were [$t = 5.17, p < .01, 2$ -tailed]. Because gender is a salient dimension on which voices differ, that listener ratings of talkers across gender indicated that language characteristics were influencing their judgments even when they were comparing male and female talkers.

3.2.4. Reported basis for judgments

All of the listeners recognized that talkers appeared more than once on the listening test. They reported that three properties provided the basis for their similarity judgments: rate (40%), pitch (40%) and tone (47%). Listeners used words such as high and deep to describe pitch. They equated tone with speaking style and even loudness. Listeners were

sensitive to the rate of speech and reported that foreign talkers speak fast.

3.3. Discussion

Listeners were generally able to identify the same talker when producing two different speech samples even in a foreign language. Female talkers were rated somewhat more variably than male talkers. Perhaps listeners find the characteristics of female voices to be complex as Murry and Singh [8] have suggested.

When both language and gender differed, listeners were able to attend to language across gender. When all female talkers were paired with Arabic males, listeners rated the Arabic females as more similar and the Latvian females as less similar to the male talkers. In a very similar pattern, when all female talkers were paired with Latvian males, listeners rated the Latvian females as more similar and the Arabic females as less similar to the male talkers.

4. GENERAL DISCUSSION

Although talker voice appeared to influence language identity judgments, listeners could identify languages at better than chance accuracy across talker gender. That is, they could ignore voice quality and attend to phonetic properties which specify a particular language, such as phoneme inventory or phoneme sequencing.

When listeners were asked to rate the similarity of voices across languages, they also seemed to be influenced by language identity. A number of phoneticians have suggested that speakers of a language share 'voice setting,' pronunciation features or articulatory habits which result from a characteristic disposition and use of the articulatory organs; suggested articulatory habits include spreadness of lips, nasality, modal pitch, and other paralinguistic parameters [2, 3, 15]. This composite of the phonetic characteristics of the language and the voice setting characteristics of talkers may be encoded as the 'sound of a language.' If language specific voice settings contribute to the 'sound of a language' these characteristics might facilitate recognition of the language across gender and also serve to make talkers of a particular language seem to have similar voice quality. Currently, however, the suggestion of 'voice setting' has to remain a hypothesis because there are no objective measures of voice similarity which could be used to provide

independent assessment of the validity of listener judgments.

Although the interaction of talker language and talker voice quality is undoubtedly complex, it appears that language identity is accessible to listeners in cross-gender judgments.

5. ACKNOWLEDGEMENTS

Preliminary versions of portions of this paper were presented at Acoustical Society of America, Ft. Lauderdale, FL, December, 2001; and the 17th International Congress of Linguists, Prague, July, 2003.

6. REFERENCES

- [1] Bond, Z.S., Stockmal, V. 2000. Distinguishing samples of spoken Korean from rhythmic and regional competitors. *Language Sciences* 24, 175-185.
- [2] Esling, J.H., Wong, R.F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL Quarterly* 17(1), 89-05.
- [3] Gelfer, M.P. 1993. A multidimensional scaling study of voice quality in females. *Phonetica* 50, 15-27.
- [4] Goggin, J., Thompson, C.P., Strube, G., Simental, L. 1991. The role of language familiarity in voice identification. *Memory and Cognition* 19(5), 448-458.
- [5] Lorch, M., Meara, P. 1989. How people listen to languages they don't know. *Language Sciences* 11(4), 343-353.
- [6] Maddieson, I., Vasilescu, I. 2002. Factors in human language identification. *Proc. of International Congress on Spoken Language Processing*, 85-88.
- [7] Mullennix, J.W., Johnson, K., Topcu-Durgun, M., Farnsworth, L. 1995. The perceptual representation of voice gender. *J. Acoust. Soc. of America* 98, 3080-3095.
- [8] Murry, T., Singh, S. 1980. Multi-dimensional analysis of male and female voices. *J. Acoust. Soc. of America* 68 (5), 1294-1300.
- [9] Muthusamy, Y.K., Cole, R.A. 1992. Automatic segmentation and identification of ten languages using telephone speech. *Proc. of the International Conference on Spoken Language Processing*, Banff, Alberta, Canada.
- [10] Palmeri, T.J., Goldinger, S.D., Pisoni, D.B. 1993. Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words. *J. of Experimental Psychology: Learning, Memory, and Cognition* 19(2), 309-328.
- [11] Pisoni, D.B. 1993. Talker normalization in speech perception. In Tohkura, Y., Vatiokiotis-Bateson, E., Sagisaka, Y. (eds.), *Speech Perception, Production, and Linguistic Structure*. Amsterdam: IOS Press.
- [12] Singh, S., Murry, T. 1978. Multi-dimensional classification of normal voice qualities. *J. Acoust. Soc. of America* 64, 81-87.
- [13] Stockmal, V., Bond, Z.S. 2002. Same talker, different language: A replication. *Proc. of International Congress on Spoken Language Processing* 77-80.
- [14] Stockmal, V., Moates, D., Bond, Z.S. 2000. Same talker, different language. *Applied Psycholinguistics* 21, 383-393.
- [15] Thornbury, S. 1993. Having a good jaw: voice-setting phonology. *ELT Journal* 47(2), 126-131.