THE DYNAMICS OF CODESWITCHING IN VOICE IDENTIFICATION

Ruth Huntley Bahr
University of South Florida, Tampa, FL, USA

ABSTRACT

Some voice quality alterations may be driven by linguistic or social forces. To be specific, these changes can result from dialectal differences, or voice patterns during specific speaking situations (i.e., casual vs. formal productions) or voice produced in a non-native language. It would seem to follow then that listeners might not recognize a voice when it is presented in a second language, as in the case of bilingualism. In this experiment, naive, monolingual listeners were asked to identify talkers in both English and Spanish in three speaking conditions: casual speech, extemporaneous speech with a monolingual speaker of the talker's second language, and reading. Results, which differed by gender, indicated that speaker identification was hampered by type of sample and language spoken. Lastly, acoustic analyses identified aspects of voice production that were altered during code-switching. Discussion will focus on how codeswitching influences voice recognition and production.

1. INTRODUCTION

Intentional changes in speaking patterns or style can influence one's ability to identify a speaker (1-3). However, not all vocal modifications can be considered intentional. Some voice quality alterations may be driven by linguistic or social forces. To be specific, these changes can result from dialectal differences, voice patterns during specific speaking situations (i.e., causal vs. formal productions) or voice produced in a non-native language. While few would argue that these types of vocal changes occur, little is known about the influence of such factors on speaker identification.

The problem of alteration in speaking style occurs frequently in criminal investigations. Typically, the relevant recordings have been made in differing environments; for example, one during the crime and another during questioning by the detectives. Hence, the possibility that the style is different between the recordings can be expected. If recognition accuracy is adversely influenced by the use of the two different speaking styles, the result may be the identification of an innocent person as the perpetrator (i.e., a false identification) or the freeing of a guilty individual (i.e., a false elimination).

2. THE PROBLEM OF LANGUAGE IN VOICE IDENTIFICATION

McGehee (4) was one of the first individuals to consider the influence of native language on voice identification. She hypothesized that a non-native English voice would be easily distinguished from a native English voice. In this case, a native German speaker using English served as the target voice in a lineup task. His voice was placed among the voices of Chinese, Greek and Russian males and another German male. McGehee found that the target voice was identified 81% of the time after a two-day delay. It was interesting to note that while the German speaker was identified at a fairly high rate, the Chinese speaker and not the other German, was most often confused with the target. This finding suggested that listeners may concentrate on voice quality more than the accent utilized. Nevertheless, the heterogeneity of the accents here may have inflated the recognition scores.

Goldstein and his colleagues (5) related the "accented" voice recognition problem to the "other race" effect noted in face recognition. In other words, faces of one's own race (or ethnic group) become more familiar more efficiently than members of a foreign group. For this experiment, they compared Taiwanese, American and African-American voices in a four-voice line. The results indicated that for short retention intervals, accented voices were no more difficult to recognize that were unaccented voices. However, when the duration of the speech sample, was reduced, performance on the accented voices declined. When the first experiment was repeated with a ten voice line using talkers speaking Spanish or accented English and a ten minute retention interval, no differences in recognition ability were noted. Again, they concluded that accented voices were no harder to recognize than unaccented if the speech sample was long enough. However, no comparison to the idenfication of English speakers was made.

Thompson (6) also considered the effect of language in voice identification. He utilized bilingual students recording messages in English, Spanish and with a strong Spanish accent. Monolingual English listeners identified "voices speaking Spanish less accurately than (the same) voices speaking English." Similar results were found by Hollien, Majewski and Doherty (7) and Doty (8) in that auditors who did not know English had more difficulty identifying talkers in English than other English-speaking participants.

Finally, Goggin and others (9) ran four experiments which assessed the "effects of language characteristics on voice identifications." They found that monolingual speakers of English and German identified speakers of their native language more accurately than speakers of a foreign language. In the third experiment, bilingual (English-Spanish) subjects identified voices equally well in both languages, as well as in the accented condition. In experiment four, they considered aspects of language familiarity that may influence recognition; in this case, phonology, lexicon and syntax. In doing so, they: 1) mixed words, preserving some syntax and all of the lexicon, 2) mixed syllables - preserving normal phonology and 3) reversed the text thereby destroying normal phonological cues. As the passages were made more remote from English, performance declined. They concluded that as the syntactic, semantic and phonologic characteristics of the message become less familiar, voice recognition declines.

These findings were upheld by the more recent work in the area of speaker identification in a foreign language (10, 11). These investigators found that knowledge of the language facilitated voice identification; however, it was not clear whether or not the typological difference between the native language and the target language was a significant factor in determining recognition performance. In this case, the Chinese listeners with knowledge of German outperformed the English and Spanish listeners who knew German. In another experiment, Schiller, Koester, and Duckworth (12) removed linguistic information from the stimulus items used for identification and found that listeners performed equally as well, regardless of native language. They believe that this finding strengthens the view that language familiarity is an important aspect of speaker identification.

The current experiment extends the findings of previous research (9-11) by examining the vocal identity of bilingual (English-Spanish) speakers more closely. These investigators reported that the cues necessary for identification are most easily

obtained when the listener understands the message. How accurate would monolingual auditors be if they were asked to identify two voices, each speaking in a different language, as belonging to one person or two different people? In a similar experiment, Bahr and Pass (13) noted that the identity of African-American talkers was obscured when they spoke in different linguistic registers. Auditor performance significantly declined in the conditions comparing African-American English with Standard English. Hence, it would seem to follow then that listeners might not recognize a voice as belonging to the same person when it is presented in a second language. Finally, anecdotal evidence would suggest that there are voice quality differences that accompany code-switching events.

The following questions are asked: (1) Do auditors recognize voices as belonging to the same individual, even when they are presented in two different languages? and (2) What acoustic changes, if any, are evidenced in code-swtiching?

3. METHOD

3.1 Participants. Eight Hispanic males and females (four each) from the Tampa Bay region provided the voices used in this experiment. Talkers were students at the University of South Florida, between the ages of 19 and 28, with no obvious speech or voice deviations and no reported hearing difficulty. Regional dialect was limited to speakers of Puerto Rican, Cuban or Panamaniam descent since these groups are known to be frequent codeswitchers. However, this factor should be of minor concern since comparisons of the same speaker, rather than across speakers, were of interest in this study. The participants currently spoke English in their living situations and were pursuing a degree at the University of South Florida. For each individual, at least one parent was Hispanic and they learned English and Spanish simultaneously as a child.

3.2. Speech recordings. Voices were gathered while the subject performed the following voice tasks: (1) talking casually with a Hispanic female, (2) speaking extemporaneously to the experimenter, (3) reading a standardized passage in Spanish, (4) reading a standardized passage in English and (5) reading sentences previously spoken in Spanish. The casual speech was gathered by placing a Hispanic female associate of the experimenter in a quiet room with the potential talker. This confederate engaged the volunteer in a conversation while the experimenter left the room. During this casual conversation, the associate engaged the participant in codeswitching from English to Spanish and back again.

The second type of speech sample was gathered while speaking to the experimenter. In an attempt to obtain a monologue from the subject, the examiner asked the talker about their family, hobbies and school activities.

The other recordings were gathered while the subject read aloud. They were asked to read two passages, one in Spanish and the other in English, as well as selected texts from the Spanish conversation with the associate.

Table 1 illustrates the possible listening pair types resulting from these recording conditions for both male and female voices. Text dependent refers to the presentation of identical text samples and text independent means that the wording of the samples is different.

Recording Cond.	Text dependent	Text independent
A-A	X	X
A-B		X
A-C	X	X
A-D		X
B-B	X	X
B-C		X
B-D		X
C-C	X	X
C-D		X
D-D	X	X

Table 1. Listing of the possible pair types to be utilized in the listening experiment. The letter A refers to casual Spanish conversation, B to casual English conversation, C to read Spanish, and D to read English.

3.3. Procedures.

3.3.1. Speech samples. Talkers were told that they were needed to make speech recordings for a research project on voice quality in Spanish. When the subject arrived, the experimenter sat the subject in a quiet room with a Hispanic associate of the researcher. The experimenter then left the room under the guise of getting materials for the experiment. The associate engaged in conversation with the talker for about ten minutes or until a sufficient amount of informal speech had been recorded on a DAT recorder. The experimental site was set up like a laboratory, so the microphone was in plain view, but the DAT and its connections were hidden. Standard topics, such as where the talker was born and raised, where they went to school and holidays, were used in an attempt to control conversational context across participants.

At the appropriate time, the experimenter reentered the room and excused the associate. The experimenter sat in a chair across from the subject and debriefed him as to the recordings made with the associate. The experimenter then asked the subject to discuss their family, where they grew up, hobbies and anything else they wished to talk about. Controls on text and background noise in this condition were similar to those utilized in the previous condition.

Finally, the talker was asked to read the first paragraph of *The Rainbow Passage* (14), a passage in Spanish, and selected statements from the earlier Spanish conversation with the associate.

3.3.2. Tape Construction. For each speaker, five phrases taken from each speaking condition were used as the target samples in a paired comparison task. These samples were digitized at 20 kHz using CSL to facilitate the construction of the paired comparison tape.

Eighty-eight pairs comparing each talker to themselves were randomly selected for this study. Twenty additional pairs were included which compared speech samples taken from two different speakers. For all pairs, within and across language condition comparisons were utilized. Recording condition and presentation order of all same and different pairs then were computer randomized. A one-second interval was placed between each sample within a pair and a five-second interval was placed between each pair. Twenty-seven of the 88 pairs were repeated as a test of listener reliability. The result was a 135-item quasi-randomized paired comparison task.

3.3.3. Listening task. Thirty-six naive listeners were selected from the university's speech-language pathology undergraduate students. These individuals had little to no experience in voice identification. After being seated in a soundproof booth, the listeners completed a brief speech listening test (15) via audiotape. No participant who scored less than 92% on this test was included in this experiment. This screening test then was followed by instructions for the paired comparison task and four voice pairs. These practice pairs were included to familiarize the listeners with the format of the test and to allow them an opportunity to devise a response set.

3.3.4. Listener judgment reliability. Thirty-one percent, or 27 voice pairs, were repeated at the end of the task to assess listener reliability. The responses from the repeated pairs for each subject were examined to determine if the subject responded in a similar fashion to each repeated pair. If a listener responded the same way to stimuli presented twice, then they would be 100% consistent in their responses. On the other hand, if a listener did not respond the same way to both presentations, then they would be 0% consistent. For this analysis, correct or incorrect responses were not of interest, only if the response was the same for both voice pair presentations. The results of this procedure revealed that listeners were 76%-87% consistent in their responses, which was acceptable for this listening task.

3.3.5. Acoustic Analyses. The speech samples were then analyzed with subroutines from SpeechStation2 (16). Monologues and reading samples from all three conditions were digitized at 20kHz and analyzed for the following: speaking fundamental frequency (SFF), speaking to pause time ratio, and speech rate. The pitch track subroutine was utilized to determine SFF for each speech sample. Since this subroutine computes the fundamental frequency (f0) for certain points in time and then averages it, the individual f0 points were saved as an Micorsoft Excel file. This information was then analyzed for the presence or absence of speech and a speaking to pause time ratio was generated. These ratios are believed to be reflective of speech timing (3). Speech rate was determined by calculating the duration of the speech sample and then dividing the number of words spoken in that time frame by the overall duration. The result was a value representing spoken words per minute. Finally, the speech samples were qualitatively analyzed for phonetic properties that remained constant across dialects.

4. RESULTS

4.1. Perceptual results. Since the reliability measure comprised the final portion of the listening task, only the responses from the first 108 pairs were used for all analyses. These answers were analyzed with respect to correct answers for the same or different voice pairs within and across each recording condition (Spanish or English) and text type (spontaneous vs. reading).

The multivariate ANOVA on the listener's responses to the same male voice pairs was significant [$F(9,153) = 2.681$; $p < 0.01$]. Post hoc testing with the LSD procedure revealed largely a speech sample effect suggesting that the vocal changes associated with codeswitching were less important than the type of spoken material being compared. It was more difficult for the listeners to identify same voice pairs when the type of spoken material varied as opposed to a change in language. Specifically, it was more difficult to compare conversational speech with read speech. This was especially true for the conversation to text pairs that also varied language,

The results for the female voices were somewhat different. The multivariate ANOVA on auditor's responses to the same voice pairs was also significant [$F(9,135) = 5.092$; $p < 0.001$]. Post hoc testing with the LSD procedure indicated that the poorer performances here were associated with more of a language effect. In other words, female voices were more likely to be confused when they were compared across languages. The speaking condition effect was less notable in these data.

It is interesting to note that listeners were less accurate in identifying same voice pairs for the female voices than they were for the male. In fact, the males were readily identifiable regardless of the language being spoken, while the females were more easily recognized in the Spanish-speaking comparisons. Since there is no compelling reason for this to be so, this finding merits further investigation.

The above results are particularly compelling in that the listeners were quite capable of judging the voices of two people as being different when in fact they were different. It is also interesting to note that some of the listeners expressed some knowledge of the Spanish language; however, it did not seem to influence their performance on the paired comparison task.

4.2. Acoustic results. Changes in SFF were noted as men and women switched from Spanish to English. Differences in prosody, as measured by speaking to pause time ratio and speaking rate were less evident in casual speech conditions. However, there were prosodic differences noted between conversation and reading. Several phonetic variations typical of Spanish remained in the English productions of these speakers. This result was to be expected (17).

5. CONCLUSIONS

In this experiment, the male voices seemed to be easier to identify than the female voices when the comparisons were made across languages. This is an interesting finding in need of more investigation since few voice identification studies have considered the female voice. One hypothesis is that female listeners, which comprised the bulk of the current auditors, perform more poorly with female voices, or perhaps there was something unique about the female voices used in this study.

The male voice comparisons seemed to be more susceptible to changes in sample type and were most likely to be in error when the sample type and language were different. On the other hand, Spanish language comparisons were more difficult for monolingual English listeners when a female voice was presented. These results support Bahr and Pass (13) who found that higher accuracy rates on a recognition task tended to bring out text effects. However, if listener accuracy was poor, then language effects begin to surface. The current findings would suggest that use of a foreign language would impede recognition performance, especially if the decisions to be made are difficult in general.

ACKNOWLEDGEMENTS

The authors would like to thank Esther McCormick for her assistance in recording the Spanish-speaking participants and running the listening task. She also appreciates the assistance of Kimberley J. Pass in the acoustic analysis phase.

REFERENCES

[1] Clifford, B. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior, 4*, 373-394.

[2] Saslove, H. & Yarmey, A. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology, 45*, 111-116.

[3] Hollien, H. (1990). *The acoustics of crime: The new science of forensic phonetics*. New York: Plenum Press.

[4] McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17*, 249-271.

[5] Goldstein, A., Knight, P., Bailis, K., & Conover, J. (1981). Recognition memory for accented and unaccented voices, *Bulletin of the Psychonomic Society*, 17, 217-220.

[6] Thompson, C. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1*, 121-131.

[7] Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress, and disguise speaking conditions. *Journal of Phonetics, 10*, 139-148.
[8] Doty, N. (1998). The influence of nationality on the accuracy of face and voice recognition. *American Journal of Psychology, 111*, 191-214.
[9] Goggin, J., Thompson, C., Strube, G., & Simental, L. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19*, 448-458.
[10] Schiller, N. O., & Koester, O. (1996). Evaluation of a foreign speaker in forensic phonetics: A report. *Forensic Linguistics, 3*, 176-185.
[11] Koester, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics, 4*, 18-28.
[12] Schiller, N. O., Koester, O., & Duckworth, M. (1997). The effect of removing linguistic information upon identifying speakers of a foreign language. *Forensic Linguistics, 4*, 1-17.
[13] Bahr, R. H., & Pass, K. J. (1995). The influence of style shifting on voice identification. *Forensic Linguistics, 3*, 24-38.
[14] Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper & Row.
[15] Griffiths, J. (1967). Rhyming minimal contrasts: A simplified diagnostic articulation test. *Journal of the Acoustical Society of America, 42*, 236-241.
[16] SpeechStation2. (1997-1998). Somerville, MA: Sensimetrics Corporation.
[17] Rogers, H. (1998). Foreign accent in voice discrimination: A case study. *Forensic Linguistics, 5*, 203-208.