On improving the perception of foreign-accented speech

Allard Jongman, Travis Wade, and Joan Sereno
Department of Linguistics, University of Kansas
E-mail: jongman@ku.edu, twade@ukans.edu, sereno@ku.edu

ABSTRACT

The present study investigates the perception of foreign-accented speech. It seeks to address the issue of whether systematic exposure to a representative sample of speech from a specific foreign accent improves comprehension of that accent. In this study, perception of Spanish-accented English was examined before and after a training regimen that involved exposure to a range or word-level Spanish-accented English productions from a number of speakers. The results show that the training failed to produce an overall advantage in participants' perception of new Spanish-accented words. While subjects showed an advantage in attending to speaker-specific information, there was an absence of learning language-specific accent characteristics.

1. INTRODUCTION

Although a person with a strong foreign accent may at first be difficult to understand, comprehension seems to improve after some initial exposure to the accent. While this impression suggests that exposure to a nonnative accent leads to improved perception, surprisingly few empirical studies have explored this claim in detail. Moreover, since English has developed into the world's *lingua franca*, foreign-accented English is poised to become the most commonly spoken language in the world, so a detailed study of the potential benefits of training on its comprehension is extremely pertinent to the facilitation of cross-cultural communication. The present study seeks to address this issue by examining the perception of Spanish-accented English before and after a training regimen previously shown to be successful in teaching language sounds and contrasts. This study is described below, following a brief discussion of relevant previous findings.

From the few previous studies dealing with perceptual learning of foreign-accented speech, it appears that the ability to perceive accented or distorted speech improves with exposure. However, clear learning effects that transfer predictably across both words and speakers have not to date been unambiguously observed. Wingstedt and Schulman [1] observed that upon training with Swedish sentences infused with an artificial 'cryptic' accent consisting of regular phonemic alternations, subjects could generalize slightly to recognize new words containing the same alternations. Similarly, Clarke [2] found that listeners become faster at recognizing words produced by a single accented speaker over the course of even very limited exposure (a few sentences) in an orthographic matching task. Furthermore, Gass and Varonis [3] found that subjects recognized accented sentences better after exposure to readings by a different speaker of the same L1 than after readings by speakers of different L1s or no readings at all, suggesting that some familiarity with the accent associated with a particular native language may affect listeners' abilities to perceive accented speech. Clarke [4] presents seemingly contradictory evidence to this claim, reporting that listeners exposed to native and accented speech over three days showed advantages attending to previously encountered speakers but not new speakers of the same L1, though this finding may have been due to the training task's emphasis of differences rather than similarities across speakers. Weil [5] offers additional mixed results, claiming that transfer of learning to other accented speakers of the same L1 may depend on the particular task used in testing. Training subjects using English word, sentence, and prose stimuli recorded by a single native speaker of Marathi, Weil found that significant generalization of learning to the speech produced by another Marathi speaker occurred only for certain sentence materials. However, it is likely that for robust learning effects to occur subjects would require exposure to multiple speakers of a given L1, and null results are therefore not surprising.

The present study seeks to address clearly the issue of whether systematic exposure to a representative sample of speech from a specific foreign accent improves comprehension of that accent. The experiment reported here employs a methodology that has proven successful in teaching foreign linguistic contrasts to American college students (e.g., Wang et al. [6]) and also in training subjects to recognize synthetic English speech (*e.g.* Greenspan et al. [7]). Results from these language-training paradigms indicate that focused training which involves exposure to sufficient levels of naturally-occurring variability in the production of sounds can be highly successful in improving learners' performance on new tokens of the same sounds. Such a high-variability training method was used to train English speakers to comprehend Spanish-accented English speech.

2. STIMULI

The training study involved exposure to word-level Spanish-accented English productions over a number of sessions. Stimuli for the study were words taken from the 20 monosyllabic 'PB-word' lists provided by Egan [8], designed to represent common English usage and to be

equal in range and degree of difficulty and phonetic content.

Productions were taken from 6 adult native speakers (3 M, 3 F) of Latin American varieties of Spanish. These speakers represented a range of English proficiency and were recruited with the primary criterion that all were judged by the experimenters to have conspicuously non-native English pronunciation. Each speaker read a unique set of three 50-word PB lists. They were first given a list of all 150 words to study and encouraged to ask questions regarding the meaning or pronunciation of any unfamiliar words. They then read the list twice, each time beginning with 5 filler items and with a short break between the two repetitions. Words were presented on a computer screen at a constant rate of 3 sec / word; words were blocked across and randomized within lists.

Recording took place in an anechoic chamber using a Fostex DAT recorder and an Electrovoice RE-20 microphone; productions were digitized at 22.05 kHz using Praat [9]. The initial production of each word by each speaker was used in training, except where articulatory or recording aberrations not related to a speaker's overall accentedness dictated that the second token should be used instead. Additionally, if a speaker produced a word once with a clearly phonemic-level mispronunciation - i.e., used a particular sound in the pronunciation of a word when another sound which was clearly in the speaker's inventory would have been more appropriate - *and* the correct (expected) sound was used in the second production, it was judged to be a one-time error and the second production was used. Productions selected for use were finally normalized for amplitude across speakers.

Pre- and post-test materials were taken from the productions of two of the speakers recorded whose recognition difficulty was judged to be representative of the set of speakers observed. Pre-test items were 50 words (one entire PB list) chosen arbitrarily from one speaker, and post-test items were another PB list from the same speaker as well as one from a new speaker. Each training session, then, consisted of 4 PB lists (one PB list produced by each of the remaining 4 speakers). This resulted in 3 unique training sessions of 200 items balanced for expected word difficulty and for number of tokens from a given speaker. No word was repeated within or across sessions. The composition and order of sessions (their combination of word lists and the day of training on which they were presented) was held constant across trainees.

3. SUBJECTS

Subjects were 30 college-age native English speakers who claimed to have little or no experience with Spanish or Spanish-accented English. 15 of the subjects were arbitrarily selected as trainees, and the remaining 15 served as controls. Subjects received course credit for their participation.

4. PROCEDURE

Testing and training sessions were administered by computer; word order was randomized, across speakers, within each session. Subjects would first hear a word over headphones, after which they were instructed to type the English word they perceived, followed by the SPACE bar. (To ensure accuracy, subjects were encouraged to type carefully rather than quickly, and reaction times were not collected).

In training sessions, subject responses were followed by feedback, in which information about the correctness of the response, the participant's typed answer, and the expected answer appeared on the screen accompanied by a bell or buzz sound (immediately after the SPACE bar terminated a response). After 1500ms, an additional repetition of the same token was played as this information remained on the screen. Finally, after another 1000ms the screen was cleared and the next token was played.

In testing, no such feedback was given, and the end of a response (SPACE bar) was simply followed by the next sound after a 1500ms pause. Typed responses homophonous with intended words within General American English were treated as correct for purposes of both feedback and scoring, while misspellings were treated as incorrect responses. All tests were administered in sound-attenuated rooms in the Kansas University Phonetics and Psycholinguistics Laboratory. Both trainee and control subjects completed pre- and post-tests, trainees on the days immediately preceding and following the three consecutive days of training, and controls in sessions 5-7 days apart.

5. RESULTS AND DISCUSSION

Pre- and post-test and overall improvement (post-test - pre-test) data are given in Table 1, averaged across the 15 control and 15 trainee subjects tested.

	pre-test		post-test		Improve-	
	score	sd	score	sd	ment	sd
controls	51%	3.19	62%	4.45	11%	3.67
trainees	53%	6.67	62%	6.08	9%	6.50

Table 1: Overall test performance of subject groups

Univariate ANOVA revealed a main effect of Test ($F=50.99$, $p<0.001$), such that subjects performed significantly better in the post-test condition; however, no effect was observed for Training ($F=0.404$, $p=0.527$), nor was there a significant Test x Training interaction ($F=0.538$, $p=0.466$). Thus, the training method employed failed to produce an overall advantage in subjects' performance on new items produced by a new and a previously encountered accented speaker.

The robust improvement across subject groups from pre- to post-test requires some explanation, as it seems unlikely that such a large difference in performance would occur by chance across word lists balanced for phonetic difficulty. Moreover, it was not the case that trainees and controls

demonstrated precisely the same pattern in post-test. Figure 1 shows the groups' scores for the two speakers encountered in post-test.

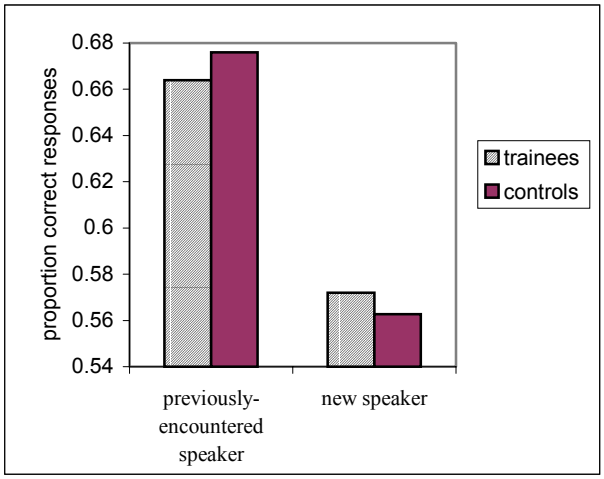

Figure 1: Post-test performance across speakers

As shown in Figure 1, words produced by the new speaker were clearly more difficult for both subject groups to comprehend ($F=33.03$, $p<.001$). Additionally, there appears to be an interaction whereby trainees show a slight advantage for the new speaker while controls prefer the previously encountered one. This effect was non-significant ($F=.357$, $p=.553$) due to the wide range of subject performance, but its relative persistence across subjects suggests the following possible scenario. The post-test improvement in accuracy for controls was likely due to a combination of (1) familiarization with the previously encountered speaker achieved during the pre-test even in the absence of a training-with-feedback mechanism, (2) similarly acquired familiarity with the task, and (3) any spurious effects of list difficulty. For trainees, factor (3) should have applied similarly, whereas effects of factors (1) and (2) might have been attenuated slightly due to forgetting the specific characteristics of the pre-test speaker's voice over the training process and its associated task. This process, then, might have been sufficient to obscure (partially) the effects of a sufficiently weak third factor for the trainees, namely the expected perceptual learning of accent-specific characteristics over training. The result, then, would represent a case of speaker-related training confusing or taking precedence over language population-specific advantages for a set of presumably related sounds.

Additional evidence for speaker-familiarity effects comes from examination of trainee performance on the four speakers encountered in training over the three training sessions, as demonstrated in Figure 2, where paired-samples t-tests demonstrated that overall performance each day superceded that of the previous day.

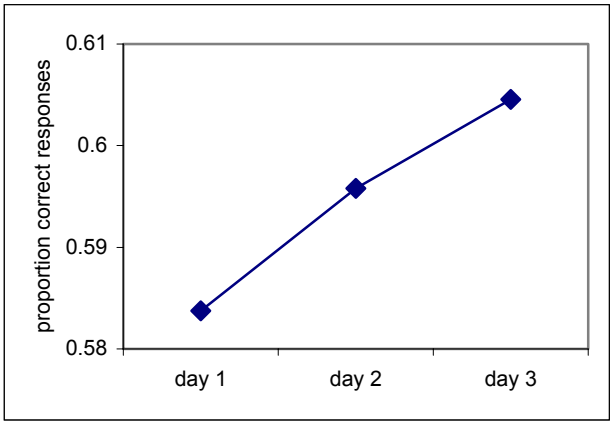

Figure 2: Session-by-session trainee performance

Thus, while training clearly gave subjects an advantage in attending to *speaker*-specific information, there was a conspicuous absence of learning for similar *accent*-level characteristics. This suggests a possible fundamental difference between the comprehension of nonnative-accented speech and that of synthetic or natively-produced foreign language sounds, considering the previous success of similar training on these sounds.

Ongoing research in our laboratory is exploring the possibility that there is an underlying difficulty related to the inherently variable nature of non-native productions due to factors including non-constant proficiency across speakers. That is, it seems likely that the sounds of non-native speech, even those from speakers of a single L1 background, are in fact not adequately described as a homogenous set of learnable deviations from a standard pronunciation. As a result, potentially generalizable features of the Spanish accent in the present study were either (1) not sufficient in quantity to produce any advantage in trainee subjects' recognitions of new accented words or (2) so easily acquired that control subjects were able to perform similarly to trainees relying only on pre-test items with the accent. To the extent that this pattern is shown to be the case, it might be suggested that training programs intended to improve comprehension of non-native speech emphasize adaptation to individual speakers—or subsets of speakers controlling for proficiency and other factors—rather than presenting subjects with the same types of variability that are effective in teaching native-produced sounds.

ACKNOWLEDGMENTS

This research was supported by a grant from the University of Kansas General Research Fund to the first author.

REFERENCES

[1] M. Wingstedt and R. Schulman, "Comprehension of foreign accents." in W. Dressler (Ed.), *Phonologica 1984: Proceedings of the 5th International Phonology Meeting* (339-45). Cambridge: Cambridge University

Press, 1987.

[2] C. Clarke, “Perceptual adjustment to foreign-accented English with short term exposure,” *Proceedings of the 7th International Conference on Spoken Language Processing* (253-56), 2002.

[3] S. Gass and E. Varonis “The effect of familiarity on the comprehensibility of nonnative speech,” *Language Learning*, 34, 65-89, 1984.

[4] C. Clarke, *Perceptual learning of foreign accented English*, Unpublished M.A. thesis, The University of Arizona, 2000.

[5] S. Weil, *Foreign accented speech: Adaptation and generalization*, Unpublished M.A thesis, The Ohio State University, 2001.

[6] Y. Wang, M. Spence, A. Jongman, and J. Sereno “Training American listeners to perceive Mandarin tones,” *Journal of the Acoustical Society of America*, 106, 3649-58, 1999.

[7] S. Greenspan, H. Nusbaum, and D. Pisoni “Perceptual learning of synthetic speech produced by rule,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (3), 421-33, 1988.

[8] J. Egan, “Articulation testing methods,” *Laryngoscope*, 58, 955-91, 1948.

[9] Praat, doing phonetics by computer. Copyright 1992-2002 by P. Boersma and D. Weenink.