Perceptual Adjustment to Foreign-Accented English
With Short Term Exposure

Constance M. Clarke

Department of Psychology
University of Arizona
Tucson, AZ 85721
clarke@u.arizona.edu

Abstract

Non-native, or 'foreign-accented,' speech deviates from native pronunciation norms, and these deviations can cause perceptual difficulty for listeners. However, there is reason to believe that non-native speech can be learned by the adult perceptual system. The present study investigated perceptual learning of foreign-accented speech with very limited experience. Sixteen auditory sentences were presented to two groups whose task was to indicate whether an orthographic word presented immediately after each sentence matched the final word of the sentence. The experimental group heard all 16 sentences of accented speech, whereas the control group heard the first four accented sentences and the last four non-accented sentences. The experimenters' response times indicated whether an orthographic word presented immediately after each sentence matched the final word of the sentence. The experimental group's response time decreased across the sixteen sentences, suggesting they rapidly adjusted to the accented speech. The experimental group was also faster for the last four accented sentences compared to the control group, indicating learning was not completely due to practice with the task.

1. Introduction

1.1 Background

Speech contains a great amount of variability from a large number of sources. How the human speech perception system extracts stable linguistic information from such a variable signal is still largely a mystery. Traditionally, the perceptual system was thought to engage in normalization when processing speech, a process of stripping away and discarding variability that does not directly specify the intended speech segment (e.g., acoustic consequences of vocal tract characteristics, phonetic context, or speaking rate [1]).

However, there is increasing evidence that variable features of speech, such as those caused by the differences between talkers, are not discarded during speech perception but are encoded and used later in perception [2], [3], [4]. For example, Nygaard and Pisoni [3] showed that familiarity with voices, gained through three days of exposure, improved intelligibility for those voices in a subsequent word identification test. This and other research demonstrate that non-linguistic variability in speech can be retained in memory and benefit perception.

1.2 Foreign accent as a source of variability

Foreign accent is a source of variability in speech that can be particularly detrimental to speech perception. Accented speech can cause misidentification of words [5], [6], and increased processing time [6]. However, accented speech is learnable by the human perceptual system. Studies have demonstrated that after laboratory training with non-native speech, listeners' performance improves on word recognition in noise [7] and echoing accuracy [8] compared to controls.

There are many aspects of the effects of experience on perception of foreign-accented speech that are still unknown; however, one is how early perceptual learning of accent can occur. Does this learning begin immediately upon initial exposure to an accented voice, or does it require hours or days of experience?

Research in the area of talker normalization suggests that very limited experience with talker-specific speech characteristics can affect perception. In a classic study, Ladefoged and Broadbent [9] manipulated the formant frequency range of a precursor sentence to examine the effect on identification of the vowel of a target word. In several cases, the same acoustic signal was identified as two different vowels depending on the formant range of the sentence context. These findings indicate that the perceptual system can make adjustments very quickly, specifically within one sentence, based on information about the talker’s speech characteristics.

Hence, we might also expect that limited experience with foreign-accented speech induces some kind of learning or adjustment by the perceptual system. This expectation matches with the subjective experience when first listening to an accented talker of initially understanding very little, but very quickly ‘tuning in’ to the accented speech and understanding it more easily.

1.3 Goals of the present study

The present study was a preliminary attempt to track perceptual learning of foreign-accented speech within the first few sentences of exposure. The methodological paradigm was patterned after a study by Dupoux and Green [10] of perceptual adjustment to highly compressed speech. In this study, listeners transcribed meaningful sentences that had been compressed to either 45% or 38% of original duration. The number of words correctly transcribed increased significantly from the first block...
of five sentences to the third block, indicating that listeners were perceptually adjusting to the speech rate.

However, in the present study transcription was not used as the measure of perceptual learning for two reasons. First, the accented speech used in the present study was highly intelligible, unlike the compressed speech in Dupoux and Green [10], and therefore would likely not elicit many word identification errors. Second, of interest in this study was not simply intelligibility but processing difficulty/efficiency. Munro and Derwing [5] showed that non-native speech can be judged as highly accented even if it is highly intelligible, and suggest that comprehensibility judgments of accented speech may be based on degree of processing difficulty. Therefore, it may be that the processing of intelligible accented speech is slowed and that this processing difficulty can be tracked if a sufficiently sensitive measure is used. The current study used reaction time (RT) as the experimental measure to assess changes in processing efficiency after limited exposure to accented speech.

1.4 Design

Processing efficiency was investigated in the present study with a cross-modal word matching paradigm. Four blocks of four auditory sentences were presented to each listener. Immediately after each sentence was presented, a word was displayed on a computer screen, and the listener's task was to judge whether that word was the same as the last word of the sentence. In order to rule out the possibility of listeners using semantic context to perform the task, only sentences that provided no predictability for the final word were used. For the experimental group, all four blocks of sentences were produced by a Spanish-accented speaker. For the control group, the first three sentence blocks were produced by a native English speaker, and only the final block was produced by the Spanish-accented speaker.

I predicted that RTs for the experimental group would decrease across the four blocks because experience with the accented voice would decrease the difficulty of processing that voice. The control group was included to ensure that the learning shown by the experimental group was not wholly due to practice effects. Hence, a secondary prediction was that the experimental group's RTs for the final sentence block would be faster on average than the control group's RTs for the final block. Because a between-subjects RT comparison was necessary, a separate measure was taken of each listener's average RT for this task. Following the experimental trials, all listeners received eight baseline trials of the same task. The sentences in these trials were produced by a non-accented speaker different from the non-accented speaker in the experimental trials. The RTs from these trials were used to normalize the experimental RTs for each subject in order to lower the subject variability for the between-subjects comparison.

2. Method

2.1 Participants

Listeners were 32 University of Arizona undergraduates (24 females, 8 males). All were native speakers of American English who were not fluent speakers of Spanish. None reported speech or hearing disorders at the time of testing. Participants were compensated with either payment or credit toward an introductory Psychology course.

2.2 Stimulus materials

Sentence materials consisted of 32 low probability sentences from the Revised Speech Perception In Noise (SPIN-R) test [11], [12]. The 16 experimental sentences (see Appendix) were recorded by a female native speaker of American English (age: 31) and by a female native speaker of Mexican Spanish (age: 45, age of English acquisition: 30), who was rated in an earlier study to have a moderate accent when speaking English. The eight practice and eight baseline sentences were recorded by another female native speaker of American English (age: 19). Amplitude was normalized to 90% of maximum for all sentences.

Because the experimental measure of processing efficiency was based on perceptual identification of the final word of each sentence (the target word), the perceptual and lexical characteristics of these words were carefully controlled. The target words from the experimental sentences were familiar, monosyllabic nouns with a mean frequency of 22.06 per million [13] and a mean of 17.75 phonemic neighbors. Both the accented and non-accented productions of the target words had been correctly identified in isolation more than 75% of the time in a prior word identification experiment.

In order to control for differences in target word duration, the accented and non-accented productions of each target word were compressed or lengthened such that they both equaled the mean duration of the two original productions. The duration of the rest of each sentence was not altered, but the mean durations for the accented and non-accented voices were similar.

For half of the experimental trials, the word presented on the computer screen was not the target word, and the correct response was no. Each of these foils was a monosyllabic English noun that was a phonemic neighbor (but not an orthographic neighbor) of the target word of the corresponding sentence. The foils differed from the target by one phoneme in either the onset (1 case), vowel (4 cases), or coda (3 cases) position. The foils were matched with the target words for mean frequency (20.00).

2.3 Procedure

The 16 experimental sentences were grouped into four blocks of four sentences each. In each block, two sentences were no trials, and two were yes trials. Four block orders were created using a Latin square design such that, across subjects, every sentence was presented in every block position. Within each listener group (experimental and control), an equal number of listeners were presented with each of the four block orders. Thus, the experimental and control groups received the same sentence materials; the only difference was that for the experimental group, all four sentence blocks were produced by the Spanish-accented speaker, while for the control group, the first three blocks were produced by the non-accented speaker and only the fourth block was produced by the Spanish-accented speaker.

1 Four of the word productions from the native English speaker were not included in the earlier word identification test, and so identification rates were not available for these words. However, in general, this non-accented speaker was highly intelligible.

2 The compression and lengthening manipulations were done using the PSOLA compression algorithm provided in the Praat wave-editing program (Paul Boersma and David Weenink, University of Amsterdam). This algorithm changes the duration of a waveform without altering its frequency properties.
Each listener sat in a quiet room in front of a computer monitor and a two-button response box. Auditory stimuli were presented at approximately 73 dB SPL(A) over headphones. All stimulus presentation and response collection was controlled by an IBM compatible computer using DMDX software. 

Each trial began with presentation of an auditory sentence. Immediately after the sentence ended, a word appeared on the computer screen and remained until the listener responded by pressing the yes button or the no button. Accuracy and reaction time (RT) feedback was provided on the computer screen after each response. If no response was given after four seconds, the trial was recorded as no response.

The order of the experiment was the eight practice trials, followed by the four experimental blocks (16 trials), and finally the eight baseline trials. A female non-accented speaker (different from the non-accented speaker in the experimental trials) produced both the practice and baseline trials, and the trials within these blocks were presented in the same order for every subject. Within each block of the experimental trials, the order of the four sentences was randomized for each subject.

3. Results

Participants were included if their error rate was less than 20% for the experimental and baseline trials (only one participant was excluded for a high error rate). Individual trials with RTs less than 200 ms or greater than 2000 ms were discarded; RTs greater than 2 standard deviations from the participant’s mean were set equal to that value. Figure 1 displays the raw mean reaction times for each sentence block (collapsed across yes and no trials) for the experimental and the control groups.

![Figure 1. Mean reaction times across four experimental blocks. Experimental group: all 4 blocks were accented; control group: blocks 1-3 were non-accented, block 4 was accented.](image)

For purposes of analysis, the mean of each listener’s baseline trials was subtracted from the means of that listener’s experimental blocks. This normalized participants’ experimental RTs according to their average RT for the task and therefore reduce between-subject variability. An independent samples t-test performed on the baseline RTs for the two listener groups showed that they were equivalent ($p > 0.5$).

To examine whether the accented speech slowed processing overall compared to non-accented speech, a comparison between the two groups was performed for the first three sentence blocks only (all accented for the experimental group, all non-accented for the control group). A 2 (group) x 3 (block) analysis of variance (ANOVA) was conducted on listeners’ normalized mean RTs, with block as a repeated measure. The analysis showed that across the three blocks, the control group (with non-accented sentences) responded faster than the experimental group (with accented sentences), $F(1,30) = 4.92, p < 0.05$.

Although reaction time, rather than accuracy, was of primary interest in this experiment, the same analysis was also performed on the error data. Parallel to the RT results, the experimental group made significantly more errors ($M = 10.24, SD = 15.5$) than the control group ($M = 1.56, SD = 6.12$), $F(1,30) = 9.87, p < 0.01$, on the first three sentence blocks. Both the accuracy and RT findings are consistent with the idea that accented speech causes processing difficulty for listeners.

With respect to the predictions of the study, the first question of interest was whether listeners who received four blocks of accented speech responded faster as they gained more experience. A one-way repeated measures ANOVA performed on the experimental participants’ mean RTs for the four sentence blocks showed that RTs decreased over the course of the experiment, $F(3,45) = 12.05, p < 0.001$. Planned comparisons revealed significant decreases from Block 1 to Block 2 and from Block 3 to Block 4 ($p < 0.01$).

The second question of interest was whether the experimental group responded faster on average on the final sentence block than did the control group. A between-subjects ANOVA revealed that the experimental group’s mean normalized RT was significantly lower for Block 4 than was the control group’s, $F(1,30) = 9.04, p < 0.01$.

Finally, it is worth noting the pattern of the means of the first two blocks for the two groups. The experimental group’s mean RT decreased by 99 ms from the first to the second sentence block, and as mentioned above, the difference was statistically significant. In contrast, the control group’s mean RT decreased by only 31.69 ms; this difference was not significant statistically. However, the control group’s mean RT decreased by only 31.69 ms, which was not significant in a post hoc comparison ($p > 0.1$). Although the Group (experimental vs. control) x Block (block 1 vs. 2) interaction did not reach significance, $F(1,30) = 2.52, p = 0.12$, this trend suggests a higher rate of initial learning for the accented voice.

4. Discussion

In this study, a group of listeners was exposed to 16 sentences produced by a Spanish-accented speaker. Previous findings of processing difficulty with foreign-accented speech [6] were replicated: RTs and error rates were higher for the first three accented blocks compared to the corresponding non-accented blocks presented to a control group. Furthermore, a decrease in this processing difficulty was shown after a small amount of exposure to the accented speech. Consistent with the experimental prediction, reaction time decreased from the beginning to the end of the 16 sentence exposure. Indeed, processing speed was shown to improve from the first to the second block of sentences, indicating that some adjustment to the accented speech occurred after only four sentences.

One possible explanation for the faster RTs across blocks is practice with the task. However, a control group who had as much experience with the task, but no prior experience with the
accented voice, were significantly slower on the last block of sentences when they were presented with the accented voice.

One possible confound in this design is that, for the control group, not only did the speech change from non-accented to accented on the fourth sentence block, but the voice changed also. Previous evidence has shown that changing voices from trial to trial during a word identification task is disruptive compared to presenting a single voice on all trials [14]. However, one interpretation of this effect is that performance is better on the single voice condition because listeners are learning the speech characteristics of the speaker. Because the current study was a preliminary investigation into adjustment to accented speech, no attempt was made to separate the effects of experience with a voice from experience with a foreign accent per se. Additional experiments are planned to control for the change in voice across groups.

Many questions remain about the adjustment of the perceptual system to foreign-accented speech. For example, how much improvement in processing efficiency is possible? The experimental group improved across the four blocks, but there was no sign of a plateau in processing time. One might speculate that RTs would have plateaued soon after the 16 sentences if more had been presented because the experimental group’s mean RT for Block 4 was equivalent to the control group’s last non-accented block (Block 3) (p > 0.5). That is, by the last block, the experimental group had overcome the processing speed deficit for the accented speech that was apparent in the first three blocks; the experimental group was processing the accented speech at the same speed as the control group processed the non-accented speech.

Further research is needed to determine what about accented speech is learned in the first moments of exposure. Some possibilities include (1) the prosodic structure of the accented speech, which could aid in word segmentation, (2) new mappings between the acoustic-phonetic input and phonological categories, and (3) new patterns of context-dependent phonetic realization. Another important question is how information about accented speech is learned. Is the improved processing efficiency with experience due to a greater dependence on lexical knowledge, or to a low level change in phonological analysis? Answers to these basic questions will provide further insight into how the human perceptual system manages this form of variability in speech.

5. References


6. Acknowledgements

This research was supported by a National Science Foundation Graduate Fellowship. I would like to thank Merrill Garrett and Ken Forster for helpful discussion of theoretical and design issues, and Christine Mahoney for assistance with statistical analyses.

7. Appendix

Experimental sentences (from the SPIN-R test [11][12]):

1. Ruth must have known about the pie.
2. She's glad Jane asked about the drain.
3. We're speaking about the toll.
4. You should not speak about the braid.
5. The old man considered the kick.
6. He is thinking about the roar.
7. I'm talking about the bench.
8. Ruth's grandmother discussed the broom.
9. I am thinking about the knife.
10. Paul has discussed the lamp.
11. Tom is considering the clock.
12. Nancy should consider the fist.
13. Paul hopes we heard about the loot.
14. Betty has considered the bark.
15. He wants to talk about the risk.
16. The old man discussed the dive.

5. References


