COMPARING INTELLIGIBILITY OF SEVERAL NON-NATIVE ACCENT CLASSES IN NOISE

Shawn A. Weil

Department of Psychology, Cognitive/Experimental Area
The Ohio State University, Columbus, OH
Sytronics, Inc., Dayton OH
weil.17@osu.edu

ABSTRACT

Increased global interaction has led to increased communication between individuals with wide ranging linguistic experience. As a consequence, the ramifications of accent in speech production need to be better understood. It is plausible that differences in speech intelligibility due to accent type affect gross intelligibility in different ways. These differences may be differentially affected by the addition of noise. The intelligibility of talkers of five language backgrounds (Ohio English, Japanese, Taiwanese Mandarin, Indian English, and Russian) was assessed using the full Speech Reception Threshold (SRT) test. These results were compared to a measure of intelligibility without noise. Global differences in intelligibility in noise between accent classes were not found, nor were changes in relative intelligibility.

1. INTRODUCTION

The ability of a native speaker of a language to interpret speech in the face of talker variability has been the subject of much investigation. An often cited example of the extremes of talker variability is foreign accented speech [1]. Breakdowns in spoken language understanding can occur when the pronunciation of a talker differs substantially from that of the listener. These breakdowns can have adverse effects when communication is critical, as in air traffic control or military applications [2]. When environmental noise (i.e., competing talkers, radio interference, cockpit noise, etc., ) is added to the auditory input, the intelligibility of the speech signal suffers further degradation.

Accented speech is not characterized by random deviations from the normative pronunciation of a particular dialect. Instead, accents are caused in part by the influence of the phonology of an individual’s native language on a second, non-native language [3]. When the second language (e.g., English) and the native language share related but slightly differing phonemes, a native pronunciation may be substituted for the standard English pronunciation. Individuals who share a common native language will produce similar substitutions in the non-native language, causing a characteristic accent. Different native language cause accents with unique, predictable deviations from non-accented speech. These differences may cause global disparities in intelligibility. All else being equal, some accents may be inherently more difficult to understand than others [4].

Noise has been shown to compound the disruptions in intelligibility due to accent in unexpected ways. An accented talker who is highly intelligible in a noiseless environment may be incomprehensible when noise is added to the speech signal. In contrast, an equally intelligible native speaker may be less adversely affected by the same masking noise [5]. It is unclear whether this pattern of results is due to the individual talker idiosyncrasies or to aspects of speech that are determined by the relationship between specific native and non-native phonologies. If the latter is true, some accents may be naturally less intelligible in noise than other accents. This possibility has implications for communication technologies in use in noisy environments.

To test this prediction, the speech Reception Threshold (SRT) [6,7,8,9] was used to assess gross differences in intelligibility between accent classes in noise. In Experiment 1, native English listeners heard talkers with diverse accent types, and talker intelligibility was assessed by finding the signal to noise ratio (SNR) that elicited perfect sentence reception during 50% of trials. In Experiment 2, the same materials were presented without noise to measure noiseless intelligibility. The results of the two measures were then correlated.

2. EXPERIMENT 1

The SRT procedure was originally designed to assess the ability of hearing impaired individuals [6]. A participant hears a sentence masked in noise, and repeats the sentence back to the experimenter. The spectrum of the noise is equal to the average spectrum of each test sentence. If all words of the sentence are correctly repeated, the next trial proceeds with a lower SNR. If any of the words are not repeated correctly, the SNR increase, making the signal more difficult to hear. In this manner, the ratio at which 50% of the sentences are correctly perceived in their entirety can be identified.

2.1. MATERIALS

The SRT task consists of ten sets of thirteen sentences. The sentences were free translations of a Dutch corpus [8]. The sentences were assigned to sets to optimally balance the phonetic inventory. An additional practice group was created using sentences from the SPIN corpus [10]. This group was only used to familiarize subjects with the procedures.
2.2. Participants and talkers

Forty-six members of the Ohio State University community participated in this experiment. All were native speakers of English, and most were raised in central Ohio. Of these, seven were paid for their participation, and 39 received course credit. Four participants were eliminated from all analysis; two because their performance was more than two standard deviations below the mean, and two because they had immediate family members who were not native speakers of English. All others reported no known hearing or speech impairments, and noted no extensive experience with accented speech.

Five accent classes were chosen for comparison. The accents were selected on the availability of qualified talkers, and for geographic diversity. The accents chosen were: Japanese, Taiwanese Mandarin, Russian, North Indian English (native Marathi and Gujarati talkers), and Ohio English. Four individuals in each accent class (two men, two women) recorded the entire corpus of SRT sentences in a sufficiently quiet room.

2.3. Procedure

The experimental session lasted one hour. The participant was presented with an initial sentence masked in noise (-6 dB SNR). This sentence was repeated with reduced noise until the participant was able to repeat the sentence without error. The next sentence was presented with additional noise (-2 dB SNR), and in subsequent sentences noise was either added or reduced as a function of the participant’s performance. The response of the participant was then transcribed by the experimenter for later analysis. After the experiment, each participant was given a short questionnaire about his language background, and was debriefed.

Each participant heard only a subset of the talkers recorded. In addition to a native practice talker, each participant was presented with a total of five talkers, one from each accent class. To assess participant reliability, a given participant heard each of these five talkers twice, once in the first half of the session, and once in the last half. The order of the accent class presentation, talker gender, and specific talker was balanced between subjects.

2.4. Results and discussion

A 2 x 5 (two presentations x five accent classes) within subjects Analysis of Variance (ANOVA) was performed. Significant main effects of accent type, \( F(4, 39) = 11.99, p < .001 \) and presentation order \( F(1, 39) = 5.29, p < .05 \) were found. The interaction of accent type and presentation was not significant \( F(4, 39) = 2.18, \text{ns} \).

Mean performance for each of the five accent classes is illustrated in Fig. 1. Error bars indicate 95% confidence intervals. Higher SRT values indicate less noise added to the signal. Planned pairwise comparisons between the accent types revealed that the amount of masking noise required to reduce the intelligibility of a native English talker to 50% (=.08 DB SNR SRT) is much greater than the noise needed to reduce the intelligibility of the other accent classes to the same threshold (3.09 dB SNR SRT). This replicates the work of [9], who found significantly lower SNR when both the talkers and listeners were native speakers.

The Russian accent elicited the highest SNR (mean = 4.17 dB SNR SRT), and the difference between the Russian and the Japanese (mean = 2.54 dB SNR SRT) and Indian (mean = 2.69 dB SNR SRT0 groups did approach significance \( p = .058 \) and \( p = .068 \) respectively). The other accent classes did not differ appreciably from each other.

3. EXPERIMENT 2

The results of Experiment 1 indicated some differences in intelligibility in noise among the accents used. However, the effects of noise on each accent class cannot be evaluated without comparing intelligibility in noise to noiseless sentence intelligibility. It is plausible that a particular accent class (e.g., Russian) may have higher than average intelligibility in ideal conditions, but lower that average intelligibility in noise. Alternatively, accent class my not be independent of disruption in noise. Noise could affect all accents similarly. Experiment 2 provides the necessary baseline to evaluate intelligibility in noise.

3.1. Method

All materials and talkers used in Experiment 1 were again used. The balancing of talkers and accents was identical. However, no masking noise was added to the sentences. Participants heard each sentence, and repeated it verbally. The experimenter transcribed the utterance.

Twenty-two members of the Ohio State University community participated in this experiment for course credit. None had participated in the first experiment. Two participants were eliminated from all analysis because they had immediate family members who were non native speakers.
of English. All others reported no known hearing or speech impairments, and noted no extensive experience with accented speech.

3.2. Results

Proportion of Errors (PE) was determined based on the number of sentences the participants failed to report correctly in their entirety. A 2 x 5 (two presentations x five accent classes) within subjects ANOVA was performed. A significant main effects of accent type, \( F(4, 39) = 10.279, p < .001 \) was found. There were no other main effects or significant interactions.

As before, performance on the English materials was superior to performance on the other accents. Mean PE (Fig. 2) also revealed higher performance for the Indian talkers compared to the other accents, which was significant when compared to the Mandarin talkers (\( p < .02 \)), and approached significance when compared to the Japanese (\( p = .076 \)) and Russian (\( p = .053 \)).

![Figure 2: Average Proportion of Errors for 5 Accent Classes](image)

Mean performance for each of the twenty talkers (four in each accent class) was collapsed over participants, and these means were correlated with mean talker SNR level from Experiment 1 (Fig. 3). This correlation was significant (\( r = .78 \)). This implies that, in general, specific talkers who are susceptible to disruption from noise also elicit higher errors without noise.

When this correlation is broken down by accent class, all correlations were high (Fig. 3, Legend). No accent class was differentially affected by the addition of noise. Instead, all of the accent classes chosen seem to be similarly affected by the addition of noise.

4. General discussion

Knowing the effects of noise on accented speech is important in situations in which critical information is being communicated by individuals with different language background. Previous studies [5] have found noise to degrade intelligibility of an accented talker to a greater degree than a similarly intelligible native speaker. The current investigation examined the possibility that this differential interaction may be due to the relationship to specific accent characteristics, leading to gross differences in intelligibility in noise.

The correlation of the results of the two experiments did not provide evidence for this hypothesis. There seemed to be no difference in the manner in which noise affects intelligibility of accented speech; it affects all accented speakers similarly. However, the current manipulation does not allow the wholesale dismissal of this hypothesis. The pattern of results in experiment 2 differed from that of Experiment 1 in several ways. Although the PE for Russian talkers in Experiment 2 was not significantly higher than the Japanese or Mandarin talkers, it had a significantly higher SRT in Experiment 1. In contrast, the PE of the Indian sentences was lower than that of the other accents, while SRT in Experiment 12 was no different. These differences may indicate a trend toward accent specific noise interactions.

In the current study, only a small number of accents were evaluated. These accents could, by chance, have similar baseline intelligibility. The talker sample used within each accent class was small, and may not have reflected talker variability within an accent class. Other factors that affect accent, such as age of second language learning, length of residence, gender, and type of instruction were not considered [3]. These possible confounds were not considered.

Averaged over many talkers with similar language background, some gross differences may emerge. However, the magnitude of these effects need to be large to warrant compensation through signal processing or other means in specific applications. Even if different accent classes do not have demonstrable difference in gross intelligibility, finer differences may emerge. Different accents may alter different aspects of second language phonology. These differences may be lost with the SRT procedure.

The SRT paradigm was originally designed to evaluate clear speech. In previous studies [7,8], great pains were made to equate the intelligibility of the talkers, optimizing inter-talker reliability to substantiate the validity of results. Accented speech is, by nature, less intelligible than unaccented speech. Similarly, the dynamic nature of the SRT task does not allow for systematic evaluation in noise at different SNRs. In future studies, other designs should be used to compare perception of different accent classes at various SNRs.
5. REFERENCES


6.0 ACKNOWLEDGEMENTS

I am indebted to Tim Anderson, Mark Pitt, Vince Schmidt, Scott Grigsby, Lisa Shoaf, Erik Tracy, and Beren Gayle Weil for their input into the preparation and execution of this study. This research was supported by the Air Force Research Laboratory (AFRL/HECA, Wright Patterson AFB, OH) through the R&SCSIL contract to Sytronics, Inc.