

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 22 (1998)
Indiana University

**Perception of “Place-Degraded Speech” by Normal-Hearing Listeners:
Some Preliminary Findings¹**

Chris Quillet, Richard A. Wright,² and David B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This work was supported by NIH-NIDCD Research DC00111 to Indiana University.

² Now at the Department of Linguistics, University of Washington, Seattle, WA.

Perception of “Place-Degraded Speech” by Normal-Hearing Listeners: Some Preliminary Findings

Abstract. This paper examines the effects of place-degraded, elliptical speech (Miller, 1956) on the speech perception performance of normal-hearing listeners. Seven pilot experiments were carried out. Each experiment used normal English sentences and place-degraded, elliptical sentences generated by a computer-controlled text-to-speech system. All stimuli were mixed in noise. Subjects were asked to complete a range of tasks, from discriminating between a pair of sentences to transcribing an entire sentence. Although floor and ceiling effects were found in several experiments, results from a number of conditions demonstrated that, as the signal-to-noise level increased, subjects’ ability to recognize and identify the elliptical speech increased while their ability to recognize and identify normal speech decreased. Applications of this methodology to hearing-impaired listeners are discussed.

Introduction

Over the last few years, there have been substantial improvements in cochlear implant designs, accompanied by an overall increase in the effectiveness of implants in adult populations. However, there remain many problems that need to be addressed if improvements in adult implant users’ speech perception are to be achieved. One of the most difficult and challenging issues has been the wide variation in CI users’ ability to recognize spoken language (Dorman, 1993). Despite comparable psychophysical data on the electrical stimulation provided by the implant, and despite the fact that adult users typically had normal speech perception abilities prior to becoming deaf, some adult CI users perform better than others in both standard audiological tests of speech perception and in informal conversational settings. The goal of this study is to explore ways to study speech perception in the more successful adult CI users. Of particular interest are strategies for coping with the degraded signal that the CI user is presented with. Coping strategies emerging from this study may be applicable to intervention work with less successful users.

Among CI users, the most common perceptual confusions involve consonants, specifically those varying in place of articulation (Tyler & Moore, 1992). This is not surprising considering that, broadly speaking, place of articulation confusions in noise are the most common type among normal-hearing listeners as well (e.g., Miller & Nicely, 1955). Place of articulation confusions have been explained most frequently in terms of the way in which place is encoded and represented in the acoustic signal. As has long been observed (e.g., Miller, 1956, *inter alia*), information about place of articulation is carried in the higher frequencies (e.g., second and third formant transitions, and release bursts and frication noise) and thus is more susceptible to distortion due to reverberation and masking. Furthermore, information about consonant place of articulation is carried in relatively transient portions of the signal, many of which are aperiodic, and is less redundantly encoded than voicing and manner information, adding to its susceptibility to loss and distortion (e.g., Wright, Pisoni, & Frisch, 1997, *inter alia*). On the other hand, information about manner and voicing is carried in the lower frequencies and in the temporal characteristics of the signal, and is generally more redundant in the signal than place of articulation information. The distortions introduced by the cochlear implant device further degrade the frequency-dependent information about consonant place of articulation.

The susceptibility of consonant place of articulation to perceptual confusion was demonstrated more than forty years ago by Miller (1956), using a methodology he called “elliptical speech.” In this procedure, trained talkers read sentences in which a consonant’s distinctive place features were replaced with another feature from the same class. That is, /p,t,k/ are switched with each other, /b,d,g/ are switched with each other, and so on. The same process was repeated for manner and voicing. This transformation results in pairs of “place-manipulated” sentences like the following example from Miller (1956):

Normal Version:	“Two plus three should equal five.”
Elliptical Version:	“Pooh kluss free soub eatwell size.”

Under normal listening conditions, the meaning of the sentence is destroyed by the elliptic speech. However, with some training, a listener can follow the elliptical speech with difficulty. When the signal is distorted through low-pass filtering, the ellipsis can no longer be detected. The low-pass filtering interferes with the manipulated distortion of the original sentence, resulting in improved identification performance. Moreover, the signal can be heard as clear speech if contextual information is introduced. Thus, under certain conditions, listeners can perceive speech as normal even when the place information has been virtually destroyed in the signal. It is possible that this sort of reorganization process may be the basis of how successful CI users overcome distortions of place information to successfully identify words in spoken language. That is, one strategy that successful CI users might employ is a much coarser analysis of the speech signal, relying less on place of articulation information than more robust features such as manner, voicing, prosodic, and contextual information, to successfully identify the words in spoken language. A first step to exploring this idea is to use more controlled experimental methods than those employed by Miller (1956) to investigate normal listeners’ abilities to deal with place of articulation manipulations. Using this methodology, we predict that any sentence that has undergone a randomization of consonant place of articulation should be virtually unintelligible in the clear, but should show increased intelligibility as distortion increases, while the unrandomized sentence should decrease in intelligibility as distortion increases. The strongest prediction is that, as distortion increases, the normal sentence and its random-manipulated derivative should be matched on identification scores until the distortion causes a catastrophic loss of information in the normal sentence. The following set of pilot experiments was a first attempt to test these predictions on groups of normal-hearing listeners.

Methods and Procedures

Stimulus Materials

The basic stimulus materials were 100 Harvard sentences (Egan, 1948; IEEE, 1969). These are English sentences made up of five key words with declarative or imperative structures. The sentences were used to generate 100 new sentences through a process of featural degradation similar to that employed by Miller (1956). The consonants in each of the five key words were replaced with a new consonant with the same voicing and manner features as the original consonants but different place features (Rabinowitz, Eddington, Delhorne & Cuneo, 1992). The replacement consonants were randomly chosen using a random number generator, with probabilities determined after the exclusion of the original consonant’s place of articulation. This was done for the voiced and voiceless stops, fricatives, and nasals. Liquids /r/ and /l/ and glides /y/ and /w/ were excluded from the randomization process.

Next, all 200 sentences, the 100 original Harvard sentences and their 100 featurally-degraded versions, were synthesized using a DECtalk DTC03 Text-to-Speech System. The speech output from the DECtalk was digitized, recorded, and saved in WAVE format using Sound Edit 16 on a Macintosh 7500/100. A 100-ms period of silence was placed immediately before and after each sentence. Next, each Harvard sentence was compared to its random-manipulated version for differences in prosody. If a

conflict in prosody was found, the random-manipulated version was re-synthesized with adjustments made to the locations of stress within the sentence. DECTalk was able to accurately recreate the stress patterns present in the original Harvard sentences in all but two of the random-manipulated sentences. This process resulted in 98 pairs of sentences. Finally, an average RMS measure was taken over the entire set of sentences. This average RMS value (70 dB) was then used to normalize the entire set of sentences.

For Experiments 1 through 6, random bit-flip (RBF) noise (Horii, House, & Hughes, 1971) was used at three different signal-to-noise levels. The RBF noise was added to the wave files using a "batch" program specifically designed to add various types of noise to sound files. This "batch" file was designed to randomly select and flip a specified percentage of the sign bits contained in a sound file. In Experiments 1 through 6, the noise was set at 0, 10, 20, and 40 percent of the samples in each file. Each of the 196 sentences (98 Harvard sentences and 98 random-manipulated sentences) was copied and saved at each of the selected RBF noise levels.

In Experiment 7, the noise source was changed from RBF to "white" noise. A Grason-Stadler 1724 Noise Generator was used to produce "white" noise at a bandwidth of 10 kHz. This noise source was then mixed with the sound files at three signal-to-noise levels, 0 dB, -5 dB, and -10 dB. Since the "white" noise produced by the noise generator was constant, the mixing of the noise with the stimuli was controlled by the Creative Labs Mixer that accompanies the Sound Blaster AWE 32. The desired signal-to-noise ratios were achieved by keeping the master volume and the wave-out volume constant while adjusting the line-in volume of the "white" noise source.

Experiment 1: Identification Transcription

Experiment 1 was designed to test subjects' ability to identify and transcribe normal and degraded sentences as a function of percent RBF noise and place feature manipulations. Upon entering the lab, subjects were told they would be listening to a list of meaningful English sentences containing familiar words. The subjects were told to listen to each sentence, up to five times if necessary, and then transcribe as accurately as possible what they heard. There were two phases to each trial, a "listening phase" during which the subject could listen to the sentence one to five times, and a "transcription phase" during which the subject was required to type into the computer the sentence he/she heard. Once in the transcription phase for a particular trial, subjects could not return to the listening phase for that trial. Prior to beginning the actual experiment, subjects were given 10 practice trials designed to familiarize them with computer-generated speech. The experiment was designed with noise level as a between-subject variable. Each subject was assigned to one of the four noise levels, with a total of five subjects per noise level. Each subject within a particular noise level received 60 trials consisting of 30 randomly-chosen normal Harvard sentences and 30 randomly-chosen place-manipulated sentences.

Subjects. Twenty subjects, ranging in age from 18 to 23, participated in this study. All subjects were enrolled in an undergraduate psychology course and received course credit for their participation. All subjects were native English speakers with no prior history of a hearing disorder.

Results. The transcription data were scored for "key" words correct, where the identification of a word in its premanipulated form was considered correct. Thus, a subject could receive a score of 0 to 5 for each sentence. Figure 1 shows the results obtained from the data analysis. In the normal Harvard sentences, as the percent RBF noise increased, the percent correct scores decreased modestly until the 40% RBF level, where a dramatic drop in transcription scores was observed. For the place-manipulated sentences, a floor effect can be seen across all three signal-to-noise levels, with scores dropping off slightly at the 40% RBF level.

 Insert Figure 1 about here.

The results of this first experiment show that the task of transcribing a place-manipulated sentence is too difficult for subjects to accomplish accurately. There was no increase in performance based on the identification scores of the random-manipulated sentences and thus our original hypothesis was not borne out under these conditions. The percent correct scores for the random-manipulated sentences were uniformly low across all signal-to-noise levels. It was clear from these data that, in order for subjects to transcribe the random-manipulated sentences with greater accuracy, they needed more context and acoustic-phonetic information about the words. To address this issue, an easier task was used to raise random-manipulated condition performance from the floor.

Experiment 2: Same/Different Task

In Experiment 2, we gave subjects a discrimination task that was designed to be much easier than the transcription task used in Experiment 1. Subjects now listened to pairs of sentences and categorized the pair as “same” or “different.” This was a much easier task than the transcription procedure used in Experiment 1 because subjects were only required to use coarse information to compare the similarity of the two stimuli and place the sentences into one of two categories. In this experiment, percent RBF noise was a between-subject variable. Thirteen subjects were assigned to each of the 0%, 20%, and 40% RBF levels, while 12 subjects were assigned to the 10% RBF level. Upon entering the laboratory, subjects were told that they would be listening to pairs of computer-generated English sentences containing familiar words. After listening to a pair of sentences, the subjects were told to label the pair as “same” if the two sentences were the same or “different” if the two sentences were not the same. Subjects’ responses were recorded using a two-button response box on which one button was labeled “SAME” and the other was labeled “DIFFERENT.” All possible combinations of normal and random-manipulated sentences were used, totaling eight different conditions:

- | | |
|----------------|---|
| 1. <i>NiNj</i> | two different normal sentences |
| 2. <i>RiRj</i> | two different random-manipulated sentences |
| 3. <i>NiNi</i> | two same normal sentences |
| 4. <i>RiRi</i> | two same random-manipulated sentences |
| 5. <i>NiRi</i> | a normal sentence and its same random-manipulated derivative |
| 6. <i>RiNi</i> | condition 5 presented in inverse order |
| 7. <i>NiRj</i> | a normal sentence and a different random-manipulated derivative |
| 8. <i>RjNi</i> | condition 7 presented in inverse order |

Each subject received an equal number of pairs across all eight conditions. Prior to beginning the actual experimental trials, subjects were asked complete the same set of practice trials used in Experiment 1 to familiarize them with computer-generated speech.

Subjects. Fifty-one subjects, ranging in age from 18 to 23, participated in this experiment. All subjects were enrolled in an undergraduate psychology course and received course credit for their participation. All subjects were native English speakers with no prior history of a hearing disorder or speech impairment.

Results. The results were scored on a basis of response probability. For example, the probability of subjects labeling a given condition as “different” and the probability of subjects labeling a given condition as “same” were calculated for each of the eight conditions across all four percent RBF levels. In

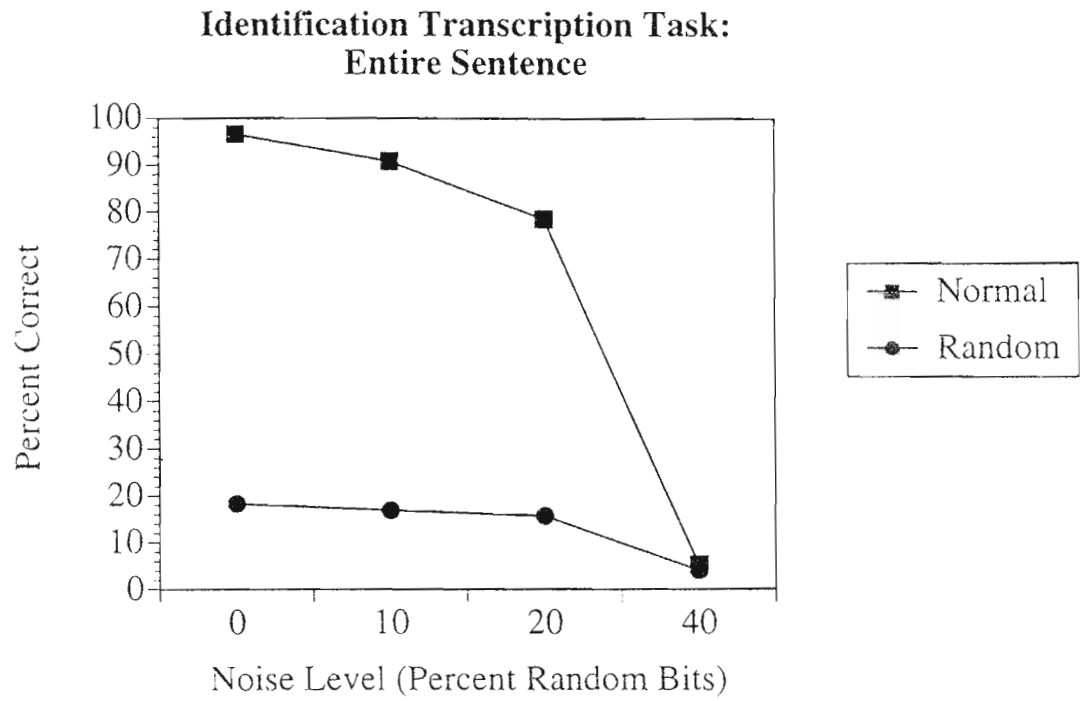


Figure 1. Percent correct recognition of key words as a function of random bit flip noise level

Condition 1, *NiNj*, subjects were able to easily distinguish between two different Harvard sentences produced by a computer. Figure 2a shows that subjects labeled these pairs of sentences as “different” with a high level of accuracy. Condition 2, *RiRj*, assessed subjects’ performance in distinguishing between two random-manipulated sentences. As shown in Figure 2b, subjects were able to accurately identify these pairs of sentences as different across all four signal-to-noise levels. Conditions 3 and 4, *NiNi* and *RiRi* respectively, tested subjects’ accuracy in labeling identical sentences presented twice. As can be seen in Figures 2c and 2d, subjects were able to correctly label these particular sentence pairs as the “same” with a high degree of accuracy. Conditions 5 and 6, *NiRi* and *RiNi*, were designed to test the subjects’ ability to discriminate between the normal and random-manipulated version of the same sentence. We found that, at the 0% RBF level, subjects were more likely to label these sentences as “different,” while at the 10%, 20%, and 40% RBF levels, subjects were more likely to label these sentences as “same.” The results of Conditions 5 and 6 are displayed in Figures 2e and 2f. Finally, Conditions 7 and 8, *NiRj* and *RjNi*, were used to assess subjects’ ability to label a different normal sentence and a different random sentence as “different.” As shown in Figures 2g and 2h, subjects were able to accurately discriminate between the two different sentences across all noise levels.

 Insert Figure 2 about here.

Conditions 1 and 2 demonstrate that subjects had no difficulty discriminating between two different normal sentences or two different random-manipulated sentences across all four percent RBF levels. Likewise, in Conditions 3 and 4, the results reveal that subjects can easily identify the same normal sentences or the same random-manipulated sentences as the same sentence across all four percent RBF levels. Similarly, in Conditions 7 and 8, subjects easily recognized the difference between a normal sentence and a different random-manipulated sentence across all four noise levels. In Conditions 5 and 6, in which subjects were asked to discriminate between the normal and the random-manipulated version of the same sentence, subjects identified a majority of the 0% RBF trials as “different,” but at the 10% RBF level and above, subjects identified a majority of the trials as the “same.” In these conditions, the likelihood of a subject responding “same” increased as a function of the signal-to-noise level. This pattern demonstrates that, as the percent RBF level increases, the normal and random-manipulated versions of the same sentence become harder to discriminate. That is, when the percent RBF level is low and the information about place is clear, subjects use this information to discriminate one sentence from the other. However, as the percent RBF level increases, the place information becomes obscured and discrimination is based on the remaining information, that is, manner and voicing. This type of pattern is exactly what is predicted.

Experiment 3: ABC Design

Experiment 3 was designed to assess subjects’ sentence identification performance without the transcription task described in Experiment 1. In this experiment, subjects were required to choose from one of three choices presented to them after each trial on their CRT monitor. Allowing subjects to view their choices allowed them to use contextual information to aid in the identification process.

Upon entering the lab, subjects were told they would be listening to a series of computer-generated English sentences containing familiar words. Their task was to listen to a sentence and then decide whether or not the sentence matched one of two normal sentences displayed on their CRT monitor or neither. Figure 3a illustrates the display shown on the subjects’ monitors. Subjects entered their responses by single-clicking with the mouse on the response of their choice. Following their response on a particular trial, subjects had to click on the “next trial” button to hear the next sentence. We used the

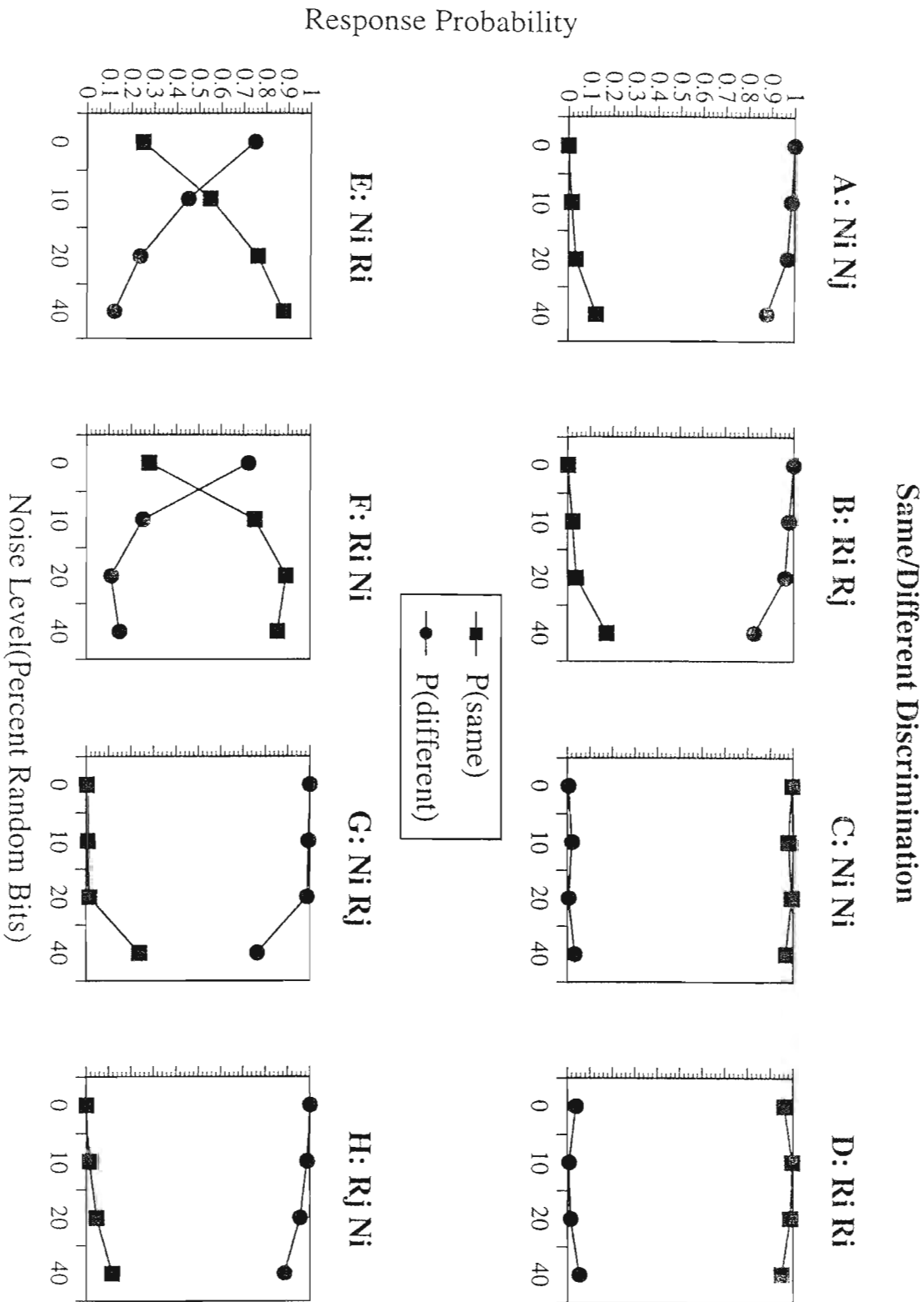


Figure 2. Probability of sentences being identified as "same" or "different" as a function of random bit flip noise level

same four percent RBF levels as Experiments 1 and 2. In this experiment, the percent RBF levels were varied within subjects. Thus, each subject received 25 percent of their trials in each percent RBF level, and 50 percent of the trials at each percent RBF level were normal Harvard sentences while 50 percent were random-manipulated sentences.

 Insert Figure 3 about here.

Subjects. Fourteen subjects, ranging in age from 20 to 24, participated in this experiment. Subjects were recruited through newspaper ads and paid \$6.00 for their participation. All subjects were native English speakers with no prior history of a hearing disorder or speech impairment.

Results. The results from Experiment 3 were scored on a percent correct basis. The results, shown in Figure 4, indicate that identification of the normal Harvard sentences remained at ceiling across the 0%, 10%, and 20% RBF levels before decreasing significantly at the 40% RBF level. In contrast, identification of the random-manipulated sentences showed small gains through the first three percent RBF levels before decreasing at the 40% RBF level.

 Insert Figure 4 about here.

While the slight increase in percent correct for the random-manipulated sentences is in the predicted direction, it is only a weak indicator that subjects are using coarse information in the identification of sentences. Furthermore, the ceiling effect for the normal sentences does not permit any conclusions to be made concerning the two conditions converging in noise. Overall, the results of Experiment 3 suggest that the ABC identification task was too easy for the normal-hearing subjects. This conclusion is supported by the fact that both the normal and random-manipulated sentences remained above the 80% correct level across the first three percent RBF levels before dropping off modestly at the 40% RBF level.

Experiment 4: Yes/No Design

Experiment 4 was designed as the complement of Experiment 3. In this experiment, subjects only saw one sentence on the CRT screen and had to choose from one of two responses, “yes” or “no.” Subjects were told they would be listening to a series of computer-generated English sentences containing familiar words. Their task was to listen to a sentence and then respond “yes” if it matched the sentence displayed on their computer monitor or “no” if it did not match the displayed sentence. Figure 5 illustrates the display on the subjects’ monitors. The subjects entered their responses by single-clicking with the mouse on the response of their choice. Following their response for a particular trial, subjects had to click on the “next trial” button to hear the next sentence. The same four percent RBF levels were used as in Experiments 1, 2, and 3. In this experiment, the noise levels were varied within subjects. Each subject received 25 percent of the trials in each of the percent RBF levels. Half of the trials at each signal-to-noise level were normal Harvard sentences and half were random-manipulated sentences.

 Insert Figure 5 about here.

ABC Interface

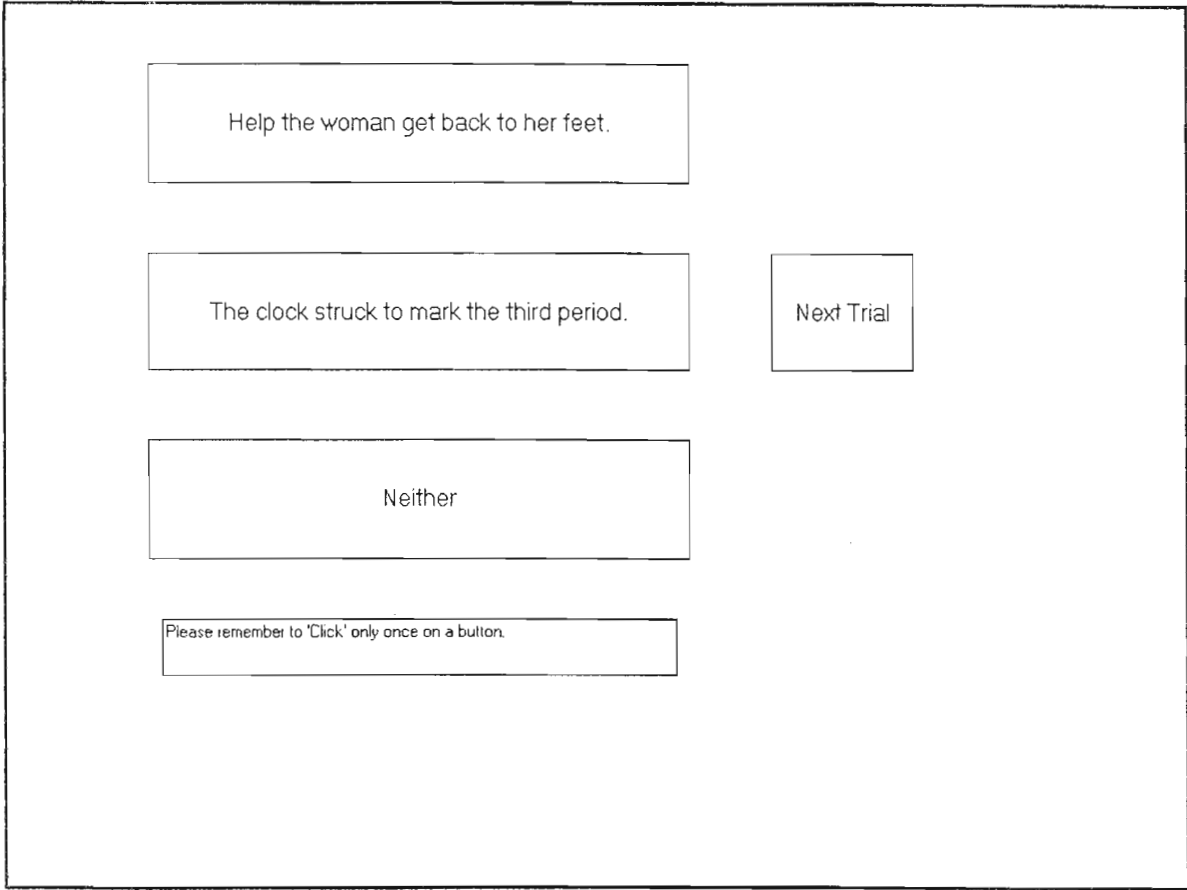


Figure 3. Graphical interface for “ABC” experimental design

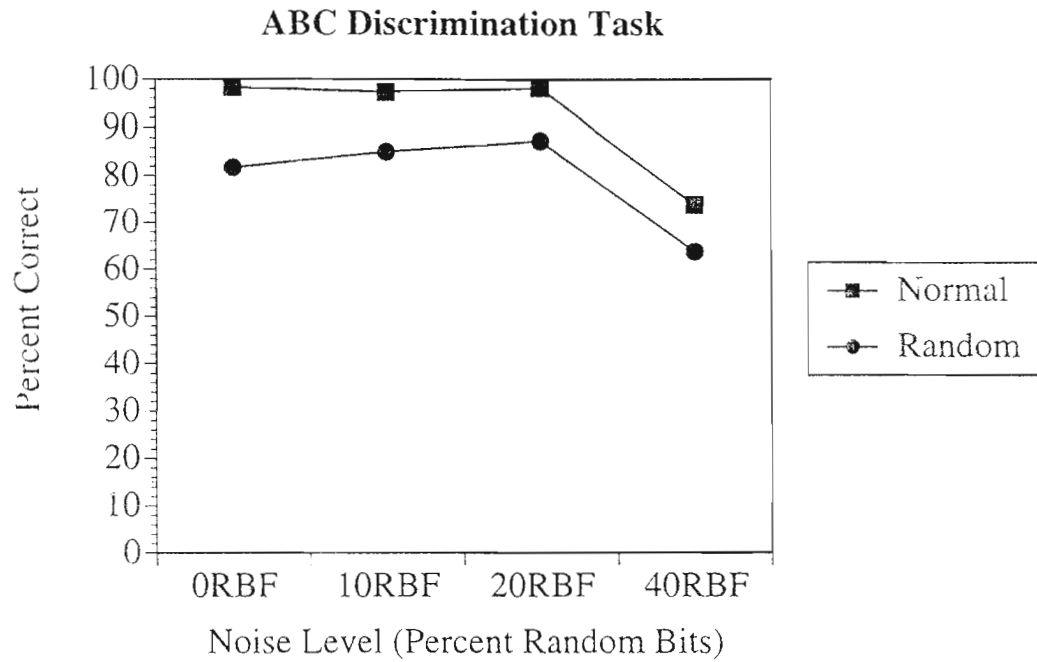


Figure 4. Percent correct identification of sentences as a function of random bit flip noise level

Yes/No Interface

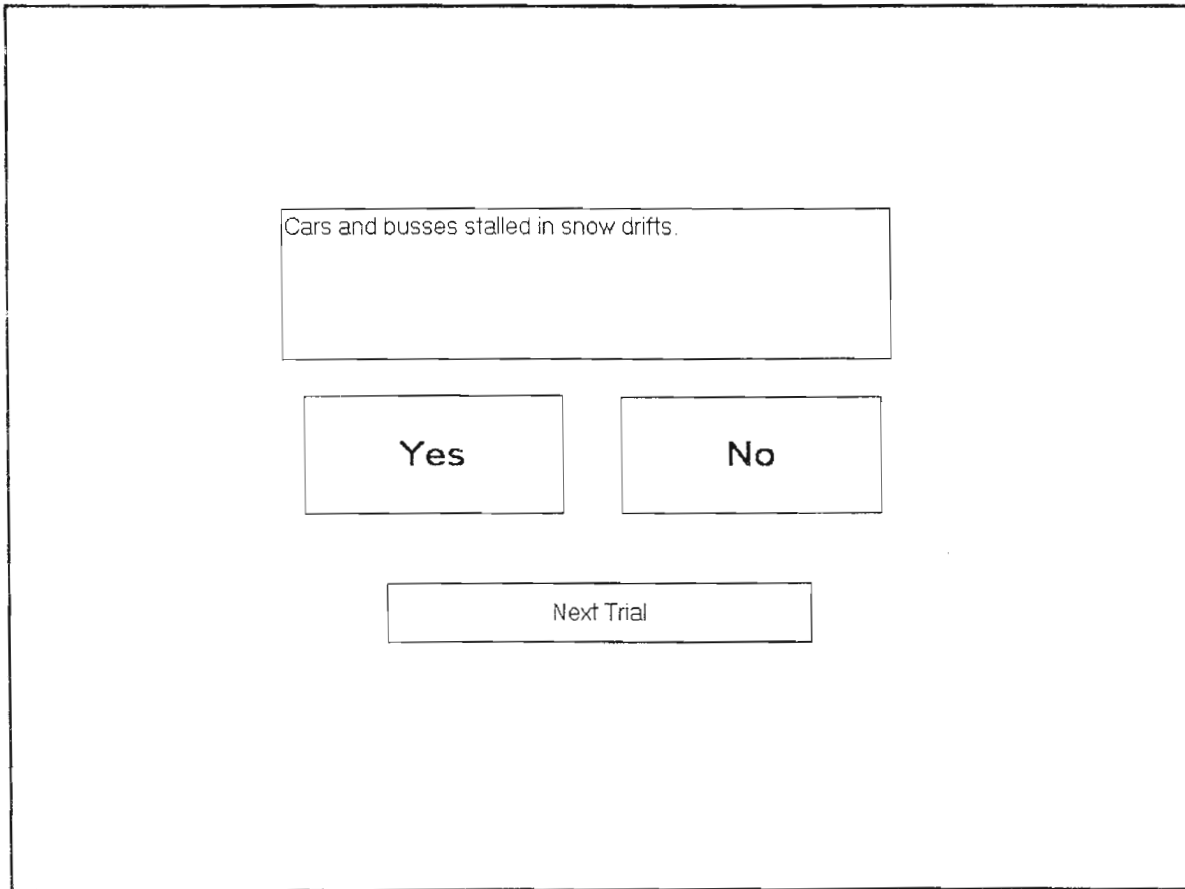


Figure 5. Graphical interface for “yes/no” experimental design

Subjects. Fifteen subjects, ranging in age from 19 to 27, participated in this experiment. Subjects were recruited through newspaper ads and paid \$6.00 for their participation. All subjects were native English speakers with no prior history of a hearing disorder.

Results. The results from Experiment 4 were scored by mean percent correct responses for each condition. The results of this experiment are illustrated in Figure 6. As shown here, the normal Harvard sentences remained at ceiling through the 0%, 10%, and 20% RBF levels, and then decreased modestly at the 40% RBF level. The random-manipulated sentences showed only small gains through the first three noise levels before decreasing at the 40% RBF level.

 Insert Figure 6 about here.

These results are similar to those obtained from Experiment 3. The percentages are slightly higher here because subjects had only two possible response choices (yes or no) whereas they had three choices (a, b, or c) in Experiment 3. Again, the hypothesis that subjects' responses to the random-manipulated sentences should become more like the responses to the normal sentences in higher signal-to-noise levels is supported by the rise in the percent correct for the random sentences. However, ceiling effects obscure the degree to which the two conditions are converging. These results indicate that a two option "yes or no" recognition task may be too easy for subjects.

Experiment 5: Transcription of Words in Isolation

This experiment was designed to illicit higher transcription scores than those found in Experiment 1. In this study, the stimuli were changed from sentences to isolated words in an attempt to reduce context effects and pull the scores down from the ceiling. From each of the original 98 normal Harvard sentences, three key words were selected along with their random-manipulated counterparts. These words were then synthesized by DECTalk, mixed with random bit-flip noise, and normalized in the same manner as the complete sentences in the previous experiments. In this experiment we used three percent RBF levels, 0%, 10%, and 20%. The 40% RBF level was excluded from this experiment because, in the previous four experiments, accuracy declined in both conditions at the 40% RBF level. In this experiment, 33 percent of the total trials were randomly assigned to each of the three percent RBF levels. Again, half of the trials in each percent RBF level were randomly assigned to normal Harvard sentences and the other half of the trials were assigned to random-manipulated sentences.

Subjects. Four subjects, ranging in age from 20 to 23, participated in this experiment. Subjects were recruited through newspaper ads and paid \$6.00 for their participation. All subjects were native English speakers with no prior history of a hearing disorder.

Upon entering the lab, subjects were told that they would be listening to a series of meaningful English words spoken in computer-generated speech. They were instructed to listen to each word, up to five times if necessary, until they felt they could accurately and confidently type the word they heard. As in Experiment 1, there were two phases to each trial, a "listening phase" during which the subject could listen to the word one to five times, and a "transcription phase" during which the subject had to type into the computer the word they just heard. Once in the transcription phase for a particular trial, subjects could not return to the listening phase for that trial.

Results. Performance was scored on a basis of percent correct words transcribed. Each transcribed word was scored as either correct or incorrect. The results of Experiment 5 are shown in

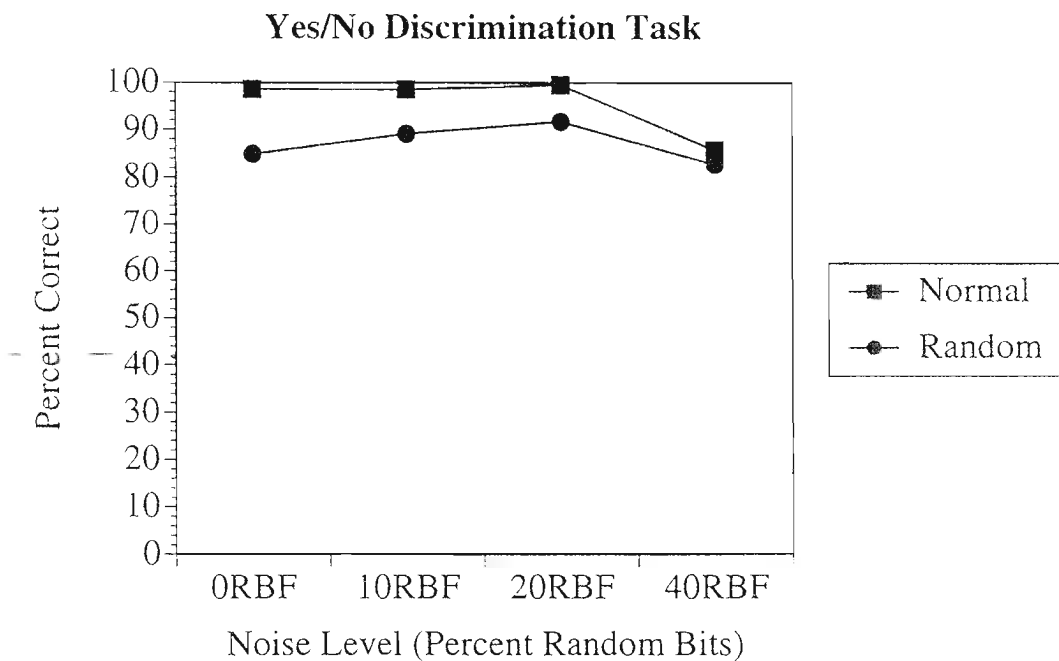


Figure 6. Percent correct identification of sentences as a function of random bit flip noise level

Figure 7. There is a steady decrease in percent correct transcription for the normal words, from 85% to 65%, across the 0%, 10%, and 20% RBF levels. There is a less dramatic increase in percent correct for the random-manipulated words, from 10% to 19%, across all three percent RBF levels. However, as in Experiment 1, the responses to the random-manipulated words appear to be flattened out by a floor effect, particularly at the 0% RBF level.

 Insert Figure 7 about here.

While the results of Experiment 5 are consistent with the predicted trends, the magnitude of the increase in the transcription of random-manipulated words is far below the expected one. The finding that the random trials remain under 20% correct across all percent RBF levels suggests that the task of transcribing words in isolation may be too difficult for subjects.

Experiment 6: Transcription of Selected Words in RBF Noise

In Experiment 5, subjects had difficulty transcribing the random-manipulated words in isolation, creating a floor effect. In Experiment 6, we attempted to raise the random-manipulated scores up from the floor by having the subjects transcribe only certain words out of a sentence. The stimuli used for this experiment were sentences mixed with RBF noise. We chose three key words from each sentence, the same key words used in Experiment 5. For each of these three key words, a blank line was substituted in a text version of the sentence. For example, in the sentence “*The show was a flop from the very start,*” we chose the words *show*, *flop*, and *start*. When the subject saw the text version of the sentence they saw: “*The _____ was a _____ from the very _____.*” See Figure 8 for an example of a subject response sheet. By showing sentences with blanks substituted for some of the words, we were able to vary the amount of context information present in the sentence.

 Insert Figure 8 about here.

Upon entering the laboratory, subjects were told that they would be listening to a series of normal English sentences containing familiar words. Subjects were then given a printed document containing the sentence stimuli with blank lines substituted for the three key words. They were instructed to listen to a sentence up to five times, and then fill in the blank lines as accurately as possible. Subjects were allowed to control the number of times a sentence was repeated by clicking with their mouse a button on their screen labeled “play the sentence.” Subjects were required to write their responses on the printed subject response sheet. Once a trial was completed, they could move on to the next trial by clicking on the button labeled “next sentence.” To insure that subjects did not lose their place on the text document, the current trial number was always displayed in the upper left corner of the monitor. See Figure 9 for a view of the CRT monitor during the experiment. The subjects’ handwritten responses were later scored by the experimenter. In this experiment, four percent RBF levels were used, 0%, 10%, 20%, and 40%. Thus, each percent RBF level was randomly assigned to 25 percent of the total trials with half of the trials for each percent RBF level randomly assigned to normal Harvard sentences and the other half assigned to random-manipulated sentences.

 Insert Figure 9 about here.

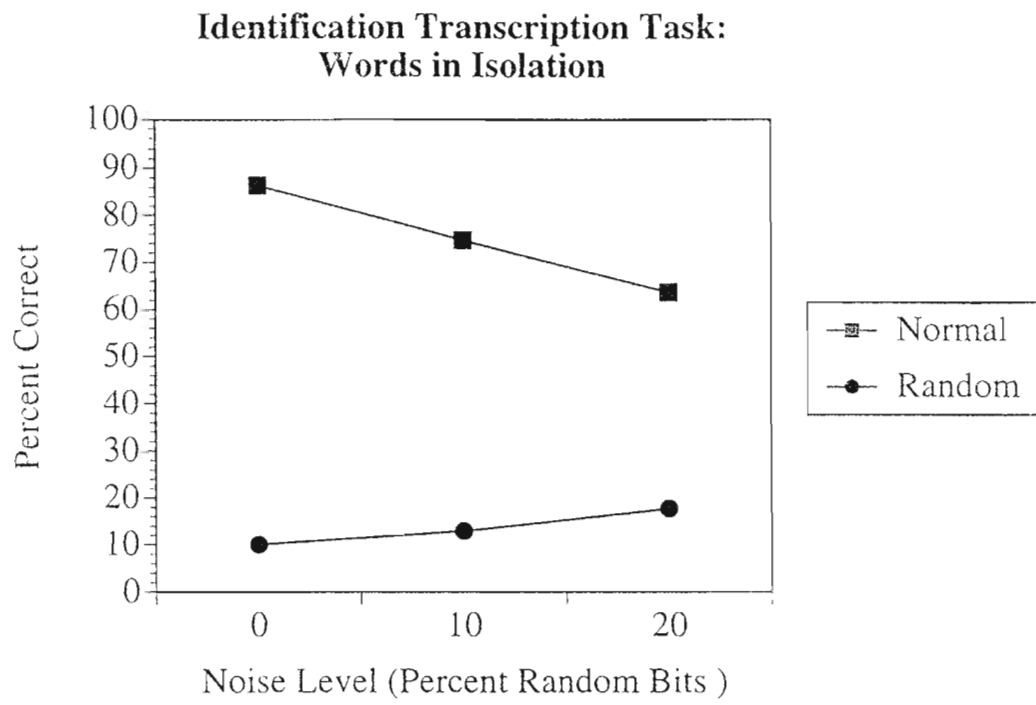


Figure 7. Percent correct recognition of words in isolation as a function of random bit flip noise level

Transcription of Selected Words: Subject Response Sheet

Output for Harvard Sentence Blanks	Name: _____
Filename: subject#.txt	
Date: 8/28/98	

1. _____ blue _____ swam in the _____.	
2. Place a _____ near the _____.	
3. It _____, rained, and _____ the same _____.	
4. The _____ to mark the third _____.	
5. The _____ cat gave _____ to _____.	
6. _____ the ball _____ and follow _____.	
7. A rod is _____ to _____ pink _____.	
8. We _____ of the _____ in the _____.	
9. There are _____ than _____ here.	
10. _____ the _____ to your left _____.	
11. The birch _____ on the _____ planks.	
12. The _____ curled around the _____.	

Figure 8. An example of the printed document used in the "Transcription of Selected Words" experiments to collect handwritten responses from the subjects

Blanks Transcription Interface

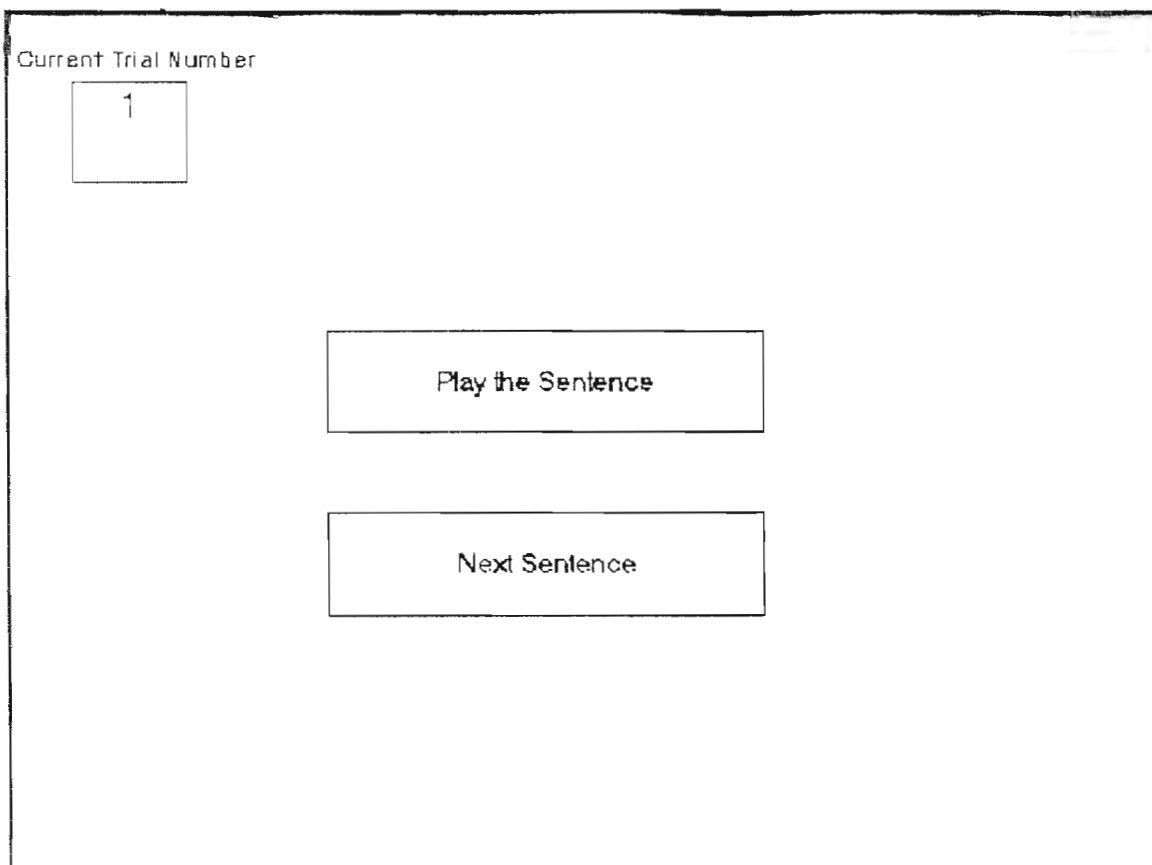


Figure 9. Graphical interface for “blanks” experimental design

Subjects. Four subjects, ranging in age from 23 to 34, participated in this experiment. Subjects were recruited through newspaper ads and paid \$6.00 for their participation. All subjects were native English speakers with no prior history of a hearing disorder.

Results. The data was scored on a percent correct basis. Each sentence was worth a total of three points, one point per word. Words were scored as correct or incorrect. A summary of the results is given in Figure 10. The normal sentence scores remained at ceiling for the 0% and 10% RBF levels, but began to show a substantial decline at the 20% RBF level. The percent correct scores for the random-manipulated sentences were higher than in Experiment 5 across all percent RBF levels. These results also showed an increasing trend from the 0% RBF level to the 10% RBF level before decreasing at the 20% and 40% RBF levels.

 Insert Figure 10 about here.

As in previous experiments, the two types of stimuli seem to converge. However, while transcription scores for the random-manipulated condition were pulled up from the floor, the normal sentence transcription scores still remained at or near ceiling for the 0%, 10%, and 20% RBF levels. It is clear from the converging trend in the data that transcription performance was improved; however, the normal sentence scores remained at ceiling. To address this issue, we changed the noise source from a random bit-flip source to a white noise source.

Experiment 7: Transcription of Selected Words in White Noise

This experiment used the same interface and design as Experiment 6. However, in all of the previous experiments, we used random bit flip noise that degrades a specified percentage of the signal. In this experiment, we switched to a white noise source, because its uniformity throughout the signal differentially masks the signal rather than degrading it. We hoped that, by altering the noise source, we would be able to accentuate the trend found in the trials using random-manipulated stimuli in Experiment 6 across all noise levels.

The white noise was mixed with the sentences at three different signal-to-noise ratios, 0 dB, -5 dB, and -10 dB. Again, subjects were told that they would be listening to a series of normal English sentences containing familiar words. They were then given a printed text document containing the blank-line version of the sentences. The subjects' task was to listen to a sentence up to five times, and then transcribe the missing words as accurately as possible. Each noise level was randomly assigned to 33 percent of the total trials, with 50 percent of the trials for each noise level randomly assigned to normal Harvard sentences and 50 percent assigned to random-manipulated sentences.

Subjects. Five subjects, ranging in age from 19 to 26, participated in this experiment. Subjects were recruited through newspaper ads and paid \$6.00 for their participation. All participants were native English speakers with no prior history of a hearing disorder.

Results. As in Experiment 6, the results were scored on the basis of percent words correct. The results are summarized in Figure 11. The identification of words in normal sentences declined markedly from the 0 dB level to the -5 dB level and again from -5dB to the -10 dB level. The identification scores of words in random-manipulated sentences showed the reverse effect, an increase in accuracy from the 0 dB level to the -5 dB level before dropping at the -10 dB level.

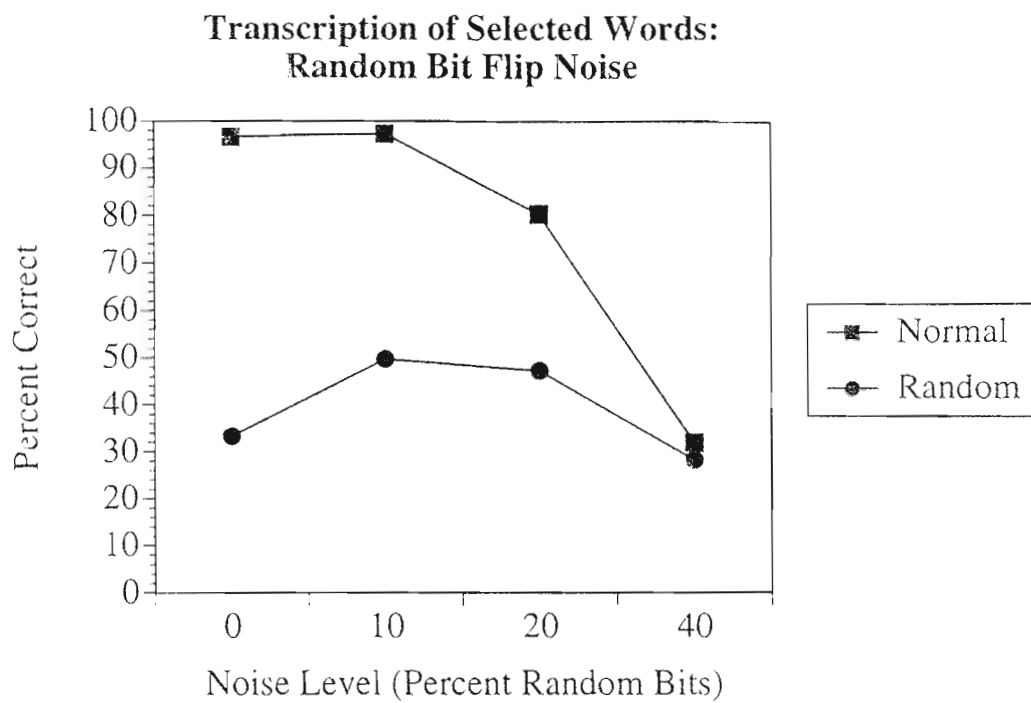


Figure 10. Percent correct recognition of key words within a sentence as a function of random bit flip noise level

 Insert Figure 11 about here.

Without the ceiling and floor effects obtained in earlier experiments, the predicted convergence occurs. This pattern demonstrates that, under the proper manipulations, naive normal-hearing listeners will use broad phonetic contrast to identify words in sentences. It is interesting to note, however, that unlike the effects reported by Miller (1956), the transcription of words from random-manipulated sentences never reached the level of accuracy obtained with the words from normal sentences.

Discussion

What does speech sound like to a post-lingually deafened adult who has a cochlear implant? This is a very difficult question to answer using the traditional behavioral assessment techniques used in clinical audiology. We know from performance on standardized speech discrimination tests that many adults with cochlear implants have great difficulty perceiving place of articulation, particularly in stop consonants. Performance on tests that measure speech feature discrimination in minimal pairs of words is often very close to chance in this population, suggesting that these contrasts are also probably difficult to perceive when the same words are presented in isolation or in sentence contexts. Despite the difficulties in perceiving place of articulation, many patients are still able to show some evidence of open-set word recognition for familiar words. They also seem to be able to comprehend large samples of fluent speech reasonably well in quiet environments. How do these patients accomplish this when they are unable to get reliable information from the speech signal about place of articulation? What does speech sound like to them through their cochlear implant? These were the basic questions that motivated the present efforts to develop a new methodology to study speech perception in this special group of listeners. One approach to this problem is to create a set of place-degraded speech signals using the methods originally developed by George Miller in his experiments on “elliptical” speech (Miller, 1962) and present these signals to patients along with the original natural speech samples. If patients treat both sets of signals as perceptually equivalent, we can infer that they “hear” natural speech with the same selective degradation and loss.

In the present set of experiments using normal-hearing listeners, we observed a trend in the predicted direction when the stimulus materials were presented in noise. That is, perception of the normal sentences became more difficult as the signal-to-noise ratio decreased while the perception of the place-degraded stimuli improved over the same range. This pattern was stronger in the same-different discrimination experiment but was also observed in the transcription experiments, particularly in the experiment using a white noise source as the masking signal instead of the original random-bit flipped noise algorithm. With the white noise source, only the weakest speech cues (i.e., place) are masked, leaving the remainder of the signal intact to support perceptual analysis and recognition of the words in the complete sentence.

While we have only used normal-hearing listeners to date in these experiments to insure that we can control and manipulate performance over a range of signal-to-noise ratios, we are planning to collect data from hearing-impaired listeners and patients with cochlear implants to see how they discriminate and transcribe these place-degraded speech signals. We are also interested in how these populations perceive natural speech signals that have not been subjected to these specific feature manipulations. Another future direction is the study of individual variation in the magnitude of these effects in both normal-hearing subjects and patients with cochlear implants. It is very likely that some of the patients with cochlear implants will be unable to do the task at all because the signals are so degraded, using these experimental manipulations, that they are unable to perceive any meaningful information from the signal. Performance

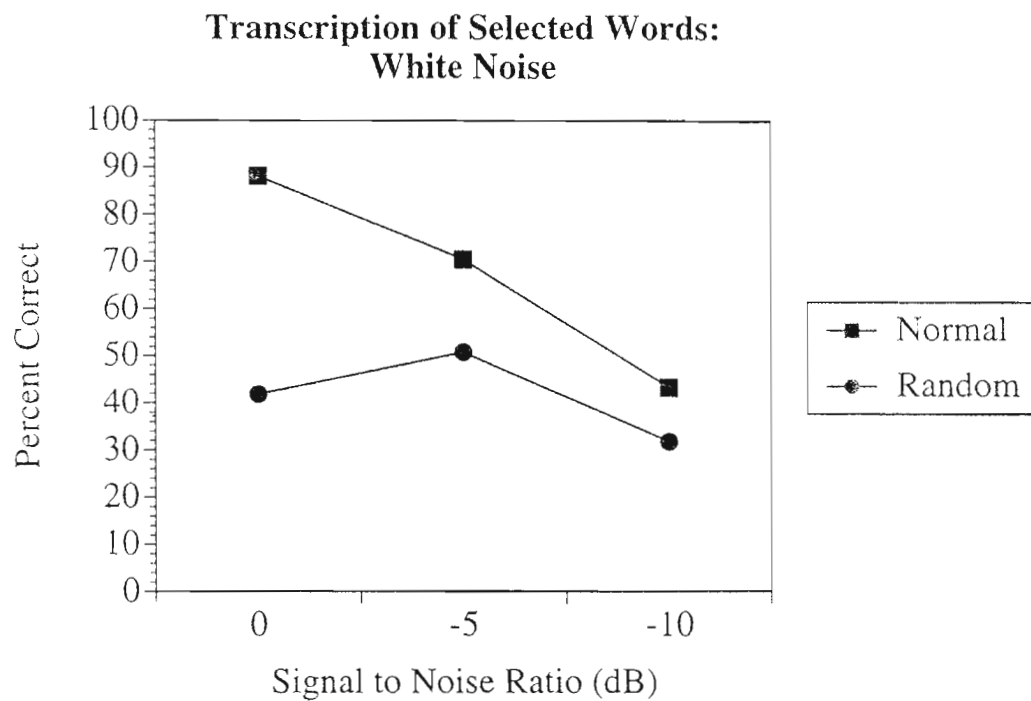


Figure 11. Percent correct recognition of key words within a sentence as a function of signal to noise ratio

with these patients may improve significantly if we switch to natural speech produced by real talkers with deliberate place errors introduced, as opposed to synthetic speech generated by rule, which is lower in overall intelligibility. If we find that patients with cochlear implants still treat the place-degraded speech samples and the original speech samples as perceptually equivalent, this would strongly imply that they are able to recognize words and understand sentences using broad phonetic categories in which certain phonetic contrasts have become “merged” into larger equivalence classes. Thus, hearing-impaired patients with cochlear implants can still recognize words and understand fluent, connected speech with these kinds of perceptually-degraded signals, but they carry this out without being able to encode fine phonetic details and subtle allophonic differences among words in the language. In linguistic terms, their lexical representations of degraded speech may therefore be even more “underspecified” than the lexical representations constructed by normal-hearing listeners. The use of “place-degraded” speech signals, like the ones originally used by George Miller in his demonstrations of “elliptical” speech, may provide some interesting new ways of investigating how these patients encode speech and how they recognize spoken words with only limited acoustic-phonetic detail. The absence of certain selective phonetic contrasts in speech may cause a “reorganization” of the perceptual space for words that differ minimally on these dimensions. This “reorganization” or “re-mapping” of listeners’ perceptual space may provide additional insights into how patients with cochlear implants are able to perceive speech as well as they do, even though they are relying on very coarse coding and impoverished acoustic-phonetic inputs at very early stages of perceptual analysis.

References

- Dorman, M.F. (1993). Speech perception by adults. In R.S. Tyler (Ed.), *Cochlear Implants Audiological Foundations* (pp. 145-190). San Diego, CA: Singular Publishing Group, Inc.
- Egan, J.P. (1948). Articulation testing methods, *Laryngoscope*, 58, 955-991.
- Horii, Y., House, A.S., and Hughes, G.W. (1971). A masking noise with speech envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America*, 49, 1849-1856.
- IEEE. (1969). IEEE recommended practice for speech quality measurements, *IEEE Report No. 297*.
- Miller, G.A. (1956). The perception of speech. In Morris Halle (Ed.), *For Roman Jakobson* (pp. 353-360). The Hague: Mouton.
- Miller, G.A. (1962). Decision units in the perception of speech. *IRE Transactions on Information Theory*, 81-83.
- Miller, G.A., and Nicely, P.E., (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-352.
- Rabinowitz, W.M., Eddington, D.K., Delhorne, L.A., and Cuneo, P.A. (1992). Relations among different measures of speech reception in subjects using a cochlear implant. *Journal of the Acoustical Society of America*, 92, 1869-1881.
- Tyler R.S. and Moore B.C.J. (1992). Consonant recognition by some of the better cochlear implant users. *Journal of the Acoustical Society of America*, 92, 3068-3077.
- Wright, R.A., Frisch, S. and Pisoni, D.B. (1997). Speech perception. In *Research on Spoken Language Processing Progress Report No. 21* (pp. 1-50). Bloomington, IN: Speech Research Laboratory, Indiana University.