# TALKER- AND TASK-SPECIFIC PERCEPTUAL LEARNING IN SPEECH PERCEPTION

*Lynne C. Nygaard and David B. Pisoni*
*Speech Research Laboratory, Indiana University, Bloomington, Indiana, U.S.A.*

## ABSTRACT

The present investigation was designed to assess the specificity of perceptual learning employed in the linguistic processing of spoken language. Two groups of subjects were trained to identify a set of talkers from sentence-length utterances. After training, one group of subjects was tested with isolated words produced by familiar or unfamiliar talkers and the other group was tested with sentence-length utterances. The results showed that the ability to identify a talker's voice from sentence-length utterances only modestly improved intelligibility of isolated words, but significantly improved the intelligibility of sentence-length utterances. Listeners appeared to focus their attention during perceptual learning on talker information that is specific to sentence-length utterances. The results suggest that task- as well as talker-specific perceptual learning occurs during the processing of spoken language.

## INTRODUCTION

The speech signal simultaneously carries information about a talker's voice and about the linguistic content of the intended message. Traditionally, the unraveling of talker and linguistic information has been characterized as a normalization process in which talker information is discarded in the listener's quest for the abstract, idealized linguistic processing units thought to underlie speech perception [1,2]. Recent studies, however, have demonstrated that the processing of voice and the processing of linguistic content are not independent. Nygaard, Sommers, and Pisoni [3] found that learning a talker's voice facilitates subsequent phonetic analysis. In their study, listeners were trained to identify talkers' voices from isolated words and were then given a word intelligibility task. Listeners who heard familiar talkers at test were better able to extract the linguistic content of isolated words than those who heard unfamiliar talkers at test. The results suggest that perceptual learning of voice can modify the linguistic processing of isolated words.

The present investigation was designed to assess the nature and extent of this kind of perceptual learning. Subjects in two experiments were trained to recognize a set of ten talkers from sentence-length utterances.

In Experiment 1, after training was completed, intelligibility was assessed using isolated words produced by familiar and unfamiliar talkers. The aim was to determine if the information learned about a talker's voice from sentences generalizes to the perception of spoken words. The assumption was that training with sentence-length utterances would focus listeners' attention at a different level of analysis than training with isolated words. It was hypothesized that because sentences contain extensive prosodic and rhythmic information in addition to the specific acoustic-phonetic implementation strategies unique to individual talkers, perceptual learning of voices from sentences would require attentional and encoding demands specific to those test materials.

In Experiment 2, after training was completed, listeners were given an intelligibility test consisting of sentence-length utterances produced by familiar and unfamiliar talkers. Two issues were addressed here. First, does specific training on sentence-length utterances generalize to similar test materials? Second, are sentence-length utterances which have higher-level semantic and syntactic constraints susceptible to the effects of familiarity with a talker's voice?

## EXPERIMENT 1

In Experiment 1, two groups of subjects learned to identify talkers' voices from sentence-length utterances over a three-day training period. The experimental group was then tested with isolated words to assess intelligibility of talkers they had been exposed to in training. The control group was tested with isolated words produced by a set of unfamiliar talkers.



### METHOD

#### Subjects

Subjects were 33 undergraduate and graduate students at Indiana University. Sixteen subjects served in the experimental condition and seventeen subjects served in the control condition. All subjects were native speakers of American English and reported no history of a speech or hearing disorder. Subjects were paid for their participation.

#### Stimulus Materials

Two sets of stimuli were used in this experiment. The sentence training stimuli consisted of 100 Harvard sentences produced by 10 male and 10 female talkers. The isolated word stimuli consisted of 100 monosyllabic words produced by 10 of the same talkers (5 male and 5 female) that produced the sentence materials. All stimuli were digitized on-line at a sampling rate of 20 kHz using 16-bit resolution. The root mean squared (RMS) amplitude levels for all stimuli were digitally equated.

#### Procedure

*Pretest Word Intelligibility.* A pretest-posttest design was used in this experiment to directly evaluate the effects of talker familiarity on word intelligibility. In both pretest and posttest, 100 isolated words produced by ten talkers (5 male and 5 female) were presented at either 80, 75, 70, or 65 dB (SPL) in continuous white noise low-pass filtered at 10 kHz and presented at 70 dB (SPL), yielding four signal-to-noise ratios: +10, +5, 0, -5. An equal number of words was presented at each of the four signal-to-noise ratios. Subjects were asked to recognize the word by typing their response on a keyboard. For subjects in the experimental condition, the words were produced by the ten talkers they heard in training. For subjects in the control condition, the talkers' voices were unfamiliar.

*Training.* Two groups of listeners completed three days of training to familiarize themselves with the voices of ten talkers. The experimental group of 16 subjects learned the voices of the same ten talkers that were used for the pre- and posttests. The control group of 17 subjects learned the voices of ten different talkers. Both groups were required to identify each talker's voice and associate that voice with one of 10 common names.

On each day of training, both groups of listeners completed three different phases. The first was a *familiarization task* in which one sentence from each talker was presented in succession. Each time a sentence was presented, the name of the talker appeared on a CRT screen in front of the listener. Subjects were asked to listen carefully to the words presented and to attend specifically to the talker's voice.

The second phase of training consisted of a *recognition task* in which subjects were asked to identify the talker who had produced each sentence. Ten sentences from each of ten talkers were presented in random order to listeners who were asked to identify each voice by pressing the appropriate button on a keyboard. On each trial, after all subjects had responded, the correct name appeared on a CRT screen.

The third phase of training was identical to the second phase except that no feedback was given.

*Posttest Word Intelligibility.* The posttest was identical to the pre-test. Subjects were asked to identify isolated words produced by familiar or unfamiliar talkers at four signal-to-noise ratios.

### RESULTS AND DISCUSSION

#### Training

All subjects showed continuous improvement over the three days of training. Both groups of subjects identified talkers consistently above chance even on the first day of training and performance rose to nearly 85% correct by the last day of training. A repeated measures analysis of variance (ANOVA) with learning and days of training as factors showed a significant main effect of day of training, $F(2,62) = 74.04$, $p<.001$, and also a significant main effect of group $F(1,31) = 20.27$, $p<.001$. The control group performed significantly better than the experimental group learning their set of talkers.

### Isolated Word Intelligibility

Figure 1 shows the difference in percent correct word identification from pretest to posttest for both the experimental and control groups averaged across signal-to-noise ratio. Although there is more improvement for subjects in the experimental condition who were hearing familiar voices at posttest than for subjects in the control condition, the effects of familiarity on word intelligibility were small ($p<.08$). A repeated measures ANOVA with signal-to-noise ratio and training group as factors showed no significant main effects or interactions.

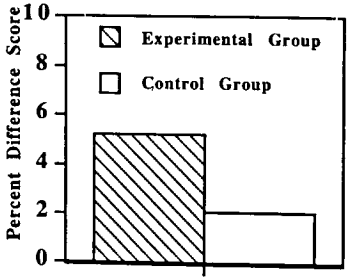


*Figure 1. Percent difference is plotted for the control and experimental groups.*

These results suggest that perceptual learning of talkers' voices from sentence -length utterances does not generalize to the perception of isolated words.

## EXPERIMENT 2

As in Experiment 1, two groups of subjects learned to identify talkers' voices from sentence-length utterances over a three-day training period. However, the experimental and control groups in this experiment were then tested with sentences produced either by talkers they had encountered in training (experimental) or by a set of unfamiliar talkers (control).

## METHOD

### Subjects

Subjects were 20 undergraduate and graduate students at Indiana University. Eleven subjects served in the experimental condition and nine subjects served in the control condition. All subjects were native speakers of American English and reported no history of a speech or hearing disorder. Subjects were paid for their participation.

### Stimulus Materials

Training and test stimuli consisted of 100 Harvard sentences produced by 10 male and 10 female talkers. All stimuli were digitized on-line at a sampling rate of 20 kHz using 16-bit resolution. The root mean squared (RMS) amplitude levels for all stimuli were digitally equated.

### Procedure

*Training.* Training was identical to that used in Experiment 1 except that subjects were trained on a set of 50 sentences rather than 100 sentences. Again, two groups of listeners completed the three days of training. The experimental group of 11 subjects learned the voices of the same ten talkers that were used for the sentence intelligibility test. The control group of 9 subjects learned the voices of ten different talkers. All other aspects of training were the same as in Experiment 1.

*Sentence Intelligibility Test.* In the sentence intelligibility test, 48 novel sentences produced by ten talkers (5 male and 5 female) were presented at either 75, 70, or 65 dB (SPL) in continuous white noise low-pass filtered at 10 kHz and presented at 70 dB (SPL), yielding three signal-to-noise ratios: +5, 0, -5. An equal number of words was presented at each of the three signal-to-noise ratios. Subjects were asked to transcribe the sentence on a sheet of paper. For subjects in the experimental condition, the sentences were produced by the ten familiar talkers they heard in training. For subjects in the control condition, the talkers' voices were unfamiliar.

## RESULTS AND DISCUSSION

### Training

All subjects showed continuous improvement over the three days of training. As in Experiment 1, both groups of subjects identified talkers consistently above chance even on the first day of training and performance rose to nearly 85% correct by the last day of training. A repeated measures analysis of variance (ANOVA) with learning and days of training as factors showed a significant main effect of day of training, $F(2,36) = 78.029$, $p<.001$, and no other



significant effects.

### Sentence Intelligibility

Subjects' performance on the sentence intelligibility task was assessed by determining the number of key words correct in each test sentence, adding up the total number of correct key words across sentences and averaging these totals across subjects. Each Harvard sentence contained 5 "key" words and the test set of 48 Harvard sentences contained 240 key words.

Figure 2 shows the total number of key words correct averaged across subjects for the experimental and control groups.

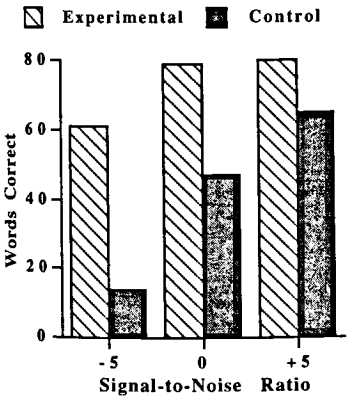


*Figure 2. Percent key words correct as a function of signal-to-noise ratio for the experimental and control groups.*

A repeated measures ANOVA with signal-to-noise ratio and training group as factors showed a significant main effect of training group, $F(1,18) = 220.378$, $p<.001$, indicating that subjects in the experimental condition who heard sentences produced by familiar talkers were able to transcribe more key words correctly across all signal-to-noise ratios than control subjects who heard sentences produced by unfamiliar talkers. A significant main effect of signal-to-noise ratio, $F(2,36) = 286.26$, $p<.001$, was also found indicating better performance at the higher signal-to-noise ratios. Finally, there was a significant interaction between training group and signal-to-noise ratio, $F(2,36) = 44.41$, $p<.001$, indicating that the effect of talker familiarity became larger as signal-to-noise ratio decreased.

These results suggest that perceptual learning of talkers' voices from sentence -length utterances facilitates the linguistic processing of sentence-length utterances produced by familiar talkers.

## GENERAL DISCUSSION

The results of our experiments suggest that perceptual learning in speech perception is both talker- and task-specific. Perceptual learning of voice transfers to linguistic processing of spoken language in a task-specific manner such that attention must be directed to learning the specific voice attributes that will be relevant at test. Our findings also show that long-term talker-specific effects on linguistic processing occur with sentence-length materials which contain higher-level semantic and syntactic constraints suggesting that talker-specific effects operate in a variety of listening situations from isolated words to sentence-length utterances.

Familiarity with a talker's voice involves long-term modification of speech and language processing. Listeners appear to retain talker-specific information about individual articulatory idiosyncrasies both at the level of acoustic-phonetic implementation and at a more global level found in sentence-length utterances.

## ACKNOWLEDGMENTS

We are grateful to Luis Hernandez for technical support, and to Lisa Burgin and Matt Peuquet for subject running. This work was supported by NIDCD Training Grant DC-00012 and by NIDCD Research Grant DC-00111 to Indiana University.

## REFERENCES

[1] Halle, M. (1985), "Speculations about the representation of words in memory," In V.A. Fromkin (Ed.), *Phonetic Linguistics* (pp. 101-104). New York: Academic Press.

[2] Joos, M.A. (1948), "Acoustic Phonetics," *Language,* vol. 24, pp. 1-136.

[3] Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994), "Speech perception as a talker-contingent process," *Psych. Sci.,* vol. 5, pp. 42-46.