euronoise

A cross-language familiar talker advantage?

Susannah Levi[a], Stephen Winters[b] and David Pisoni[c]

[a]University of Michigan, 440 Lorch Hall, 611 Tappan Street, Ann Arbor, MI 48109, USA
[b]University of Calgary, Social Sciences Bldg. Room 820, Department of Linguistics, Calgary, AB, Canada T2N 1N4
[c]Indiana University - Speech Research Laboratory, 1101 E 10th St., Dept. of Psychology, Bloomington, IN 47401, USA
svlevi@umich.edu

Previous research has shown that familiar talkers are more intelligible than unfamiliar talkers, although the cause of this processing advantage remains unknown. In the current study, we tested the source of this familiar talker advantage by manipulating the type of talker information available in the signal. Two groups of listeners were trained to identify the voices of German-English bilingual talkers; one group learned the voices from German stimuli and the other from English stimuli. After three days of training, all listeners performed a word recognition task in English. Consistent with previous findings, English-trained listeners made fewer errors with stimuli from trained talkers than untrained talkers. German-trained listeners, however, showed no familiar talker advantage, suggesting that listeners must have knowledge of relevant, language-specific information to elicit the familiar talker advantage.

1 Introduction

A speech signal transmits both the linguistic content of an utterance and indexical information about the talker, such as their gender, age, sociolinguistic background, and personal identity [1]. A growing body of literature shows that these two types of information interact in speech processing; linguistic processing is faster and/or more accurate in single-talker conditions compared to multiple-talker conditions [2, 4, 5], in same-talker compared to different-talker conditions [8, 9], and with acoustically similar talkers compared to acoustically different talkers [3]. Listeners also identify words produced by familiar talkers more accurately than words produced by unfamiliar talkers [6, 7], a finding we will refer to as the "familiar talker advantage".

In this study we investigated the factors responsible for the familiar talker advantage by controlling the type of indexical information available to listeners. Previous research has shown that indexical information can be both language-dependent and language-independent. Language-dependent indexical properties are those talker cues which are tied to the linguistic information encoded in the speech signal, such as dialectal and idiolectal articulations. In contrast, language-independent indexical properties are cues to talker identity that are not tied to linguistic information and are present when a talker speaks different languages. It has been shown that listeners can use language-independent indexical cues to both identify and discriminate between bilingual talkers across two different languages [10].

The familiar talker advantage has only been demonstrated in studies which have trained listeners to identify talkers speaking in one language and then tested the listeners' identification of words in the same language. The familiar talker advantage may therefore be the result of either language-dependent or language-independent talker information facilitating the process of word recognition because both types of indexical information are available to listeners. In this study, we tested whether knowledge of language-independent indexical properties alone was sufficient to induce the familiar talker advantage. We trained listeners to identify talkers speaking in either English or German, and then tested those listeners' ability to identify spoken English words by both familiar and unfamiliar talkers. Winters, Levi, and Pisoni [10] have shown that native English listeners trained to identify talkers from German stimuli rely primarily on language-independent cues to talker identity, whereas native English listeners trained to identify talkers from English stimuli depend more heavily on language-dependent indexical properties of voices to identify talkers. If listeners require knowledge of language-dependent indexical properties to exhibit the familiar talker advantage, then German-trained listeners should not show improvement in English word recognition for familiar talkers, while English-trained listeners should show this effect.

2 Experiment

2.1 Methods

2.1.1 Stimulus materials

Ten female German L1/English L2 speakers living in Bloomington, IN, were recorded in a sound-attenuated booth at the Speech Research Laboratory at Indiana University and paid $20 for their participation. Speech samples were recorded using a SHURE SM98 head-mounted microphone. A single repetition of 360 English and 360 German monosyllabic CVC words was produced by each speaker. All stimuli were normalized to a uniform RMS amplitude. Based on data collected in a pilot word recognition study, talkers were divided into two groups ("Group 1 talkers", "Group 2 talkers") of approximately equal intelligibility. Listeners identified an average of 45.8% whole words correct over four signal-to-noise ratios for Group 2 talkers and 42.7% words correct for Group 1 talkers.

2.1.2 Participants

Thirty-two listeners participated in the German-training condition and 32 in the English-training condition. All listeners were native speakers of American English attending Indiana University. None reported any knowledge of German. All were between the ages of 18-25, reported no history of speech or hearing impairments, and were paid $10/hour for their participation. Half of the listeners in each language training condition were trained on Group 1 Talkers ("Group 1 Listeners") and half on Group 2 Talkers ("Group 2 Listeners").

2.1.3 Procedure

Participants were trained to identify either the Group 1 Talkers or the Group 2 Talkers by name in six training sessions spanning three days. Each talker was associated with a common female name in both English and German.

In each training session, listeners completed two training blocks followed by a testing block. Each training block began with a brief familiarization phase in which listeners heard a set of words produced by each of the five talkers.

After familiarization, listeners completed a recognition task in which they identified the talkers from individual words. During these recognition phases, listeners heard five different tokens from each of the five talkers, presented twice in random order, and received feedback by seeing the correct talker's name while hearing the stimulus token again. After two training blocks, listeners completed a testing phase similar to the recognition task but without feedback. The testing phase consisted of 10 words produced once by each of the five speakers, in random order. Participants completed two training sessions per day for three days.

On the fourth day of the experiment, listeners completed a word recognition task in which they heard monosyllabic CVC English words and typed their responses. Stimuli in the word recognition test were presented to listeners at four different signal-to-noise ratios (SNR): Clear (no noise added), +10, +5, and 0 dB SNR. One quarter of the stimuli were presented at each SNR. Typed responses to the word recognition test were coded for whole word accuracy and for the number of correct phonemes per response (0-3).
German-trained listeners heard all 360 English words during the word recognition task, while English-trained listeners heard only 180 words, none of which had been presented during training. One third of the stimuli were spoken by Group 1 talkers, one third by Group 2 talkers, and one third by five native speakers of English. Data from the native English talkers will not be reported.

2.2 Results

2.2.1 Training

An Analysis of Variance (ANOVA) was conducted on the response data from the test phases of the six training sessions and assessed the effects that Training Session (1, 2, 3, 4, 5, 6) and Training Language (English, German) had on the percentage of talkers correctly identified in each testing phase. The ANOVA revealed a significant main effect of training session ($F(5,62) = 84.34$; $p < .001$), but no effect of training language, nor an interaction between training session and training language. The main effect of training session indicated that talker identification accuracy improved across the six training sessions. The lack of a main effect of training language or an interaction between training language and training session suggests that listeners learned the talkers to the same degree and at the same rate regardless of the training language. These results are illustrated in Figure 1.

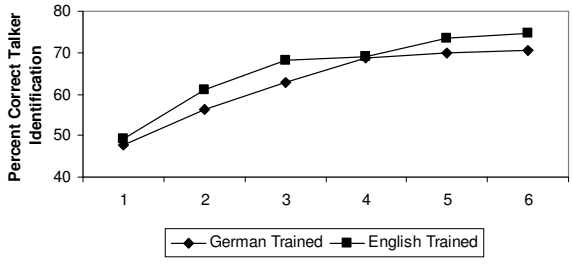

Fig.1 Talker identification accuracy during the six training sessions for both German-trained and English-trained listeners. Two training sessions were completed on each day of training.

2.3 Word recognition

Listeners were divided into "good learners" and "poor learners" based on the criterion used by Nygaard & Pisoni [6], who found that listeners who did not reach 70% accuracy in talker identification did not show the familiar talker advantage. In the German-training condition, 9 out of 16 Group 1 Listeners and 7 out of 16 Group 2 Listeners were classified as good learners. In the English-training condition, 8 out of 16 Group 1 Listeners and 12 out of 16 Group 2 Listeners were good learners.

English-trained listeners. Separate ANOVAs for good and poor learners were run on the whole word correct data with Talker Group (Group 1 talkers, Group 2 talkers) and SNR (clear, +10, +5, 0 dB SNR) as within-subjects factors and Listener Group (trained on Group 1 talkers, trained on Group 2 talkers) as a between-subjects factor. For the good English-trained learners, a main effect of SNR was found ($F(3,54) = 2.15$, $p < .001$), indicating that listeners performed worse at more difficult SNRs. In addition to this main effect, the Talker Group by Listener Group by SNR interaction also reached significance ($F(3,54) = 2.92$, $p = .041$) and the Talker Group by Listener Group interaction approached significance ($F(1,18) = 3.63$, $p = .071$). This latter crossover interaction indicates that good English-trained learners perceived more whole words correct for trained talkers than for untrained talkers. This result is displayed in Figure 2 where the outer bars for the good learners (Group 1 talkers matched with Group 1 listeners and Group 2 talkers matched with Group 2 listeners) are higher than the inner bars. The significant three-way interaction results from different patterns of responses at each SNR, driven mostly by a large benefit of talker familiarity at the +5 dB SNR, and less benefit at the other SNRs. For the poor English-trained learners, only a main effect of SNR was found ($F(3,30) = 339.7$, $p < .001$).

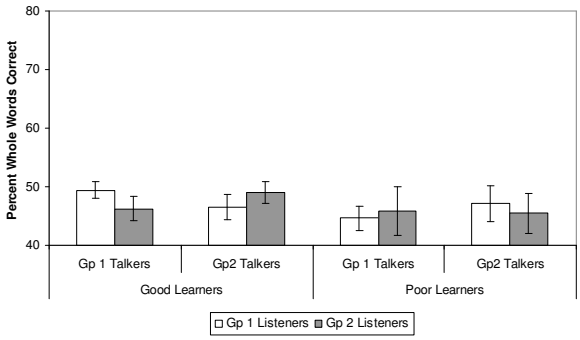

Fig.2 Percent whole words correct for English-trained listeners.

Similar results were found for the number of phonemes correctly identified during word recognition (Figure 3). For the good English-trained learners, a main effect of SNR was found ($F(3,54) = 8.65$, $p < .001$). In addition, the Talker Group by SNR interaction ($F(3,54) = 3.12$, $p = .032$) and the Talker Group by Listener Group interaction ($F(1,18) = 8.67$, $p = 0.008$) were significant. As with the whole word correct data, the Talker Group by Listener Group interaction indicated that good learners perceived more phonemes correct when listening to familiar talkers than to unfamiliar talkers. Paired-samples t-tests of the Talker Group by SNR interaction revealed that in the clear condition, listeners perceived more phonemes correct for

the Group 2 talkers than for the Group 1 talkers ($p = .036$), likely reflecting the inherent differences between the two talker groups; no differences in talker intelligibility were found for the other three SNRs. For the poor learners, the main effect of SNR reached significance ($F(3,30) = 5.32$, $p < .001$), as did the Talker Group by Listener Group by SNR interaction ($F(3,30) = 3.60$, $p = .025$). Further examination of this three-way interaction revealed that in the clear listening condition, poor learners actually perceived more phonemes correct for untrained talkers than for trained talkers.

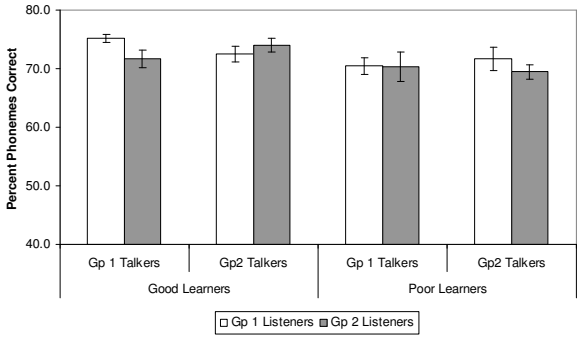

Fig.3 Percent phonemes correct for English-trained listeners.

German-trained listeners. The pattern of results for the German-trained listeners differs from that of the results obtained for the English-trained listeners. For the good German-trained learners, only the main effect of SNR reached significance ($F(3,42) = 263.6$, $p < .001$). No other main effects or interactions reached significance. For the poor learners, main effects of SNR ($F(3,42) = 299.3$, $p < .001$) and Talker Group ($F(1,14) = 5.932$, $p=0.029$) were found. The main effect of SNR again shows the benefit of increased SNR. The main effect of Talker Group indicates that the poor learners found Group 2 talkers more intelligible than Group 1 talkers. This difference in average intelligibility for the poor learners likely reflects the inherent intelligibility differences in the two groups of talkers.

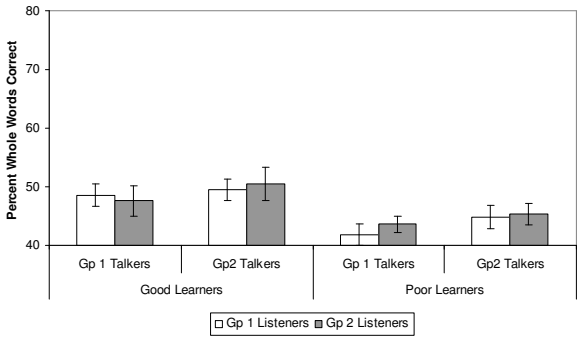

Fig.4 Percent whole words correct for German-trained listeners.

Similar results were obtained for the number of phonemes correctly identified during word recognition (Figure 5). For the good German-trained learners, only the main effect of SNR reached significance ($F(3,42) = 363.9$, $p < .001$). For the poor learners, main effects for SNR ($F(3,42) = 332.8$, $p < .001$) and Talker Group were found ($F(1,14) = 6.10$, $p=0.027$). As with the whole word correct data, poor German-trained learners perceived more phonemes correctly for Group 2 talkers than Group 1 talkers.

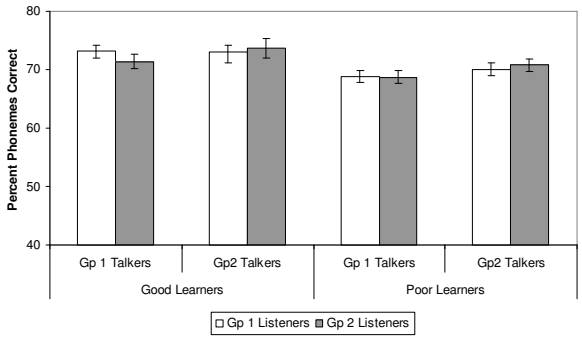

Fig.5 Percent phonemes correct for German-trained listeners.

3 Discussion and conclusion

The present study demonstrates that linguistic processing is only improved for familiar talkers when listeners have learned language-dependent indexical cues to talker identity. Consistent with previous findings, English-trained listeners identified words and phonemes produced by familiar talkers more accurately than words and phonemes produced by unfamiliar talkers. In contrast, German-trained listeners exhibited no differences in word identification accuracy between familiar and unfamiliar talkers. In other words, even though the ability to identify a talker transfers from German to English [10], knowledge these language-independent indexical properties does not facilitate linguistic processing in English. Thus, it appears that a listener must know language-dependent indexical information about a talker's voice to exhibit a familiar talker advantage on a linguistic processing task.

We attribute these findings to the type of indexical information available in the two training conditions. Listeners trained on English stimuli learned language-dependent, English-specific indexical properties, which were also present in the English stimuli used during word recognition. In contrast, listeners trained on German stimuli learned only language-independent indexical properties, without access to English-specific indexical information. These listeners could not, therefore, draw upon their familiarity with the talkers to help them perform the English word recognition task, because this knowledge contained no language-dependent information relevant to the identification of English words.

The results of the current study provide additional evidence that linguistic processing is performed in a talker-contingent manner, but also demonstrates that listeners must have knowledge of language-dependent talker information to facilitate linguistic processing in a word recognition task. The absence of a familiar talker advantage for the German-trained listeners suggests that the familiar talker advantage is not due to knowing a voice *per se* or to being able to identify different talkers, but rather to knowing how a talker produces linguistically significant contrasts in a specific language.

Acknowledgments

This work was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant DC-00012 and NIH-NIDCD Research Grant R01 DC-00111). We would like to thank Melissa Troyer for her help with data collection.

References

[1] Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine Publishing Company.

[2] Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 152-162.

[3] Magnuson, J. S. & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, **33**. 391-409.

[4] Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics,* **47***,* 379-390.

[5] Mullennix, J. W., Pisoni, D. B, & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**. 365-378.

[6] Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, **60**, 355-376.

[7] Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.

[8] Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **19***,* 309-328.

[9] Schacter, D. L. & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 915-930.

[10] Winters, S. J., Levi, S. V., & Pisoni, D. B. (in press). Identification and discrimination of bilingual takers across languages. *Journal of the Acoustical Society of America*.