

RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 28 (2007)

Indiana University

A Cross-Language Familiar Talker Advantage?¹

Susannah V. Levi,² Stephen J. Winters,³ and David B. Pisoni

*Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405*

¹ 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.

² Currently at the University of Michigan.

³ Currently at the University of Alberta.

A Cross-Language Familiar Talker Advantage?

Abstract. Previous research has shown that familiar talkers are more intelligible than unfamiliar talkers. 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 five 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 found the speech of trained talkers to be more intelligible than untrained talkers, as measured by whole words and phonemes correct. German-trained listeners, however, showed no familiar talker advantage, suggesting that listeners must have knowledge of talker-specific, linguistically relevant information to elicit the familiar talker advantage.

Introduction

The speech waveform conveys both indexical and linguistic information. Indexical information includes details about the talker, such as gender, age, sociolinguistic background, and personal identity (Abercrombie, 1967). Linguistic information forms the content of the utterance. Although listeners can selectively attend to either one of these two dimensions, a growing body of literature shows that these two types of information interact in speech processing. Linguistic experience has been shown to affect indexical processing; listeners are better able to identify and discriminate talkers in their native language (Goggin, Thompson, Strube, & Simental, 1991; Thompson, 1987) or in a second language (Köster & Schiller, 1997; Schiller & Köster, 1996; Schlichting & Sullivan, 1997; Sullivan & Schlichting, 2000) than in an unfamiliar language. Similarly, aspects of the indexical dimension can affect linguistic processing; linguistic processing is faster and/or more accurate in single-talker conditions compared to multiple-talker conditions (e.g., Goldinger, Pisoni, & Logan, 1991; Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989), in same-talker compared to different-talker conditions (Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992), with acoustically similar talkers compared to acoustically different talkers (Magnuson & Nusbaum, 2007), and with familiar compared to unfamiliar talkers (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). In this study we investigated the factors that are responsible for this latter effect – the familiar talker advantage – by training listeners to learn the voices of bilingual talkers either in English or in an unknown language. This manipulation allowed us to control the type of indexical information that listeners received and test whether language-specific indexical information is necessary to elicit the familiar talker advantage.

What do listeners know about familiar talkers that they do not know about unfamiliar talkers that facilitates linguistic processing? When listening to a talker, listeners have access to both language-specific indexical information and language-independent indexical information. Language-specific indexical properties are tied to the linguistic information encoded in the speech signal, such as dialectal and idiolectal articulations of the talker. Because these indexical properties are associated with linguistic contrasts in the language, they are not available as cues to talker identity in other languages. In contrast, language-independent indexical properties are cues to talker identity that are available across different languages. These properties include size and shape of the vocal tract, gender, and age.

The existence of language-independent indexical properties has been demonstrated recently in a study using cross-language talker identification and discrimination tasks (Winters, Levi, & Pisoni, submitted). Monolingual English listeners learned to identify the voices of bilinguals speaking in either

English or German and were later tested on their ability to generalize this knowledge to the other language. Both groups of listeners were able to identify talkers above chance in the untrained language, even when it was an unknown language. In a second experiment that measured cross-language talker discrimination, untrained English listeners were asked to judge whether two words were spoken by the same or different talker in both matched language (both English or both German) and mismatched (one stimulus in English, one in German) conditions. Results from this experiment revealed that language-independent indexical properties exist and are sufficient to support accurate talker discrimination across different languages. From these results Winters et al. concluded that some aspects of a talker's identity must be retained when speaking different languages (i.e. language-independent indexical properties) and that listeners are able to reliably use those acoustic attributes of speech to perform voice identification and discrimination tasks.

When listeners know both types of indexical information about a set of talkers, they show facilitation in linguistic processing tasks. Nygaard, Sommers, and Pisoni (1994) trained native English listeners to learn the voices of ten unfamiliar talkers speaking in English and then had them complete a speech intelligibility task with words from both familiar (i.e., trained) and novel talkers mixed with noise. Results revealed a familiar talker advantage with greater word recognition accuracy for familiar talkers compared to unfamiliar talkers. Because listeners learned the novel voices using English words, they were able to learn both language-specific and language-independent indexical information about the talkers and could use this knowledge to facilitate linguistic perception. Other research confirms that listeners learn and store subphonemic, talker-specific, linguistically relevant articulations and use this knowledge in a talker-contingent manner when performing phoneme categorization tasks (Allen & Miller, 2004; Eisner & McQueen, 2005; Kraljic & Samuel, 2005). In these studies, listeners learned to associate a potentially acoustically ambiguous segment with a particular phoneme (t/d, f/s, or s/j) and then generalized this knowledge about the category boundary to new stimuli spoken by the same talker but not by different talkers.

While knowledge of both language-specific and language-independent indexical properties can elicit the familiar talker advantage, it remains unclear whether knowledge of language-specific indexical properties is necessary for talker-contingent effects to be observed. To examine this issue, we controlled the type of indexical information available in the signal by familiarizing listeners with talkers in different languages. Two groups of monolingual English listeners were trained on the voices of L1 German/L2 English bilingual talkers. One group learned the voices from German stimuli, while the other group learned the voices from English stimuli produced by the same set of talkers. After training, listeners performed a word recognition task in English with both familiar talkers and unfamiliar talkers. With this manipulation, we were able to isolate the type of indexical information that listeners were able to learn from the talkers. Listeners trained with German stimuli could only learn talker-general, language-independent characteristics (and possibly some German-specific characteristics), whereas listeners trained with English stimuli not only learned language-independent indexical properties of the talker's voice but also acquired detailed, English-specific properties. If listeners require knowledge of language-specific indexical properties to exhibit a familiar talker advantage, then listeners in the German training condition should not show a facilitation during English word recognition for familiar talkers.

Experiment

Methods

Stimulus Materials. Twelve female German L1/English L2 speakers living in Bloomington, IN, were recorded in a sound-attenuated IAC booth at the Speech Research Laboratory at Indiana University.

Speech samples were recorded using a SHURE SM98 head-mounted unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20,000 Hz. Utterances were digitized into 16-bit stereo recordings via Tucker-Davis Technologies System II hardware at 22,050 Hz and saved directly to an IBM-PC. A single repetition of 360 English and 360 German words was produced by each speaker. Each word was of the form consonant-vowel-consonant (CVC) and was selected from the CELEX English and German databases (Baayen, Piepenbrock, & Gulikers, 1995). Stimulus materials were presented visually to speakers in random order and blocked by language. (See Levi, Winters, & Pisoni, 2007 for additional details about the recording methods.) The recording session lasted approximately one hour per language. The silent portions before and after each stimulus were removed by hand using Praat sound editing software, and the resulting tokens were normalized to a uniform RMS amplitude of 66.5 dB. German was selected as the second language in the experiment because it has a sufficient number of CVC words with the same syllabic structure as the English words and because uniformly calculated frequency counts for both the English and German words were available in the CELEX database.

The bilingual speakers were given the option of recording the materials in two sessions, but all speakers elected to record all stimuli in a single recording session. Bilingual speakers were paid \$10 per hour for their time. Two speakers were eliminated (speech disorder, N=1; greater age difference: N=1), yielding 10 bilingual speakers. Based on data collected in a pilot word-recognition study, talkers were divided into two groups (“Group 1 talkers,” “Group 2 talkers”) of roughly equal intelligibility. Average intelligibility scores, as well as other demographic data, are provided in Table 1.

Talker Group	Speaker	Age of Acquisition	Years of English	Length of Residence	Fluency	Intelligibility
1	F3	10	14	1	5	49.0
	F4	13	13	3	4.5	43.8
	F7	9	12	1	5	33.5
	F9	9	16	2	4	48.4
	F10	13	11	5	5	38.5
	Mean (SD)	10.8 (2.1)	13.2 (1.9)	2.4 (1.7)	4.7 (.4)	42.7 (6.6)
2	F2	12	9	1	4	37.1
	F5	10	14	5	3	41.1
	F8	13	16	4	5	54.8
	F11	--	--	2	4.5	41.3
	F12	7	26	5	5	54.8
	Mean (SD)	10.5 (2.6)	16.2 (7.1)	3.4 (1.8)	4.3 (.8)	45.8 (8.3)

Table 1. Demographic variables for the bilingual speakers. “Years of English” refers to the number of years speakers have been learning/using English (current age – age of acquisition). “Fluency” is a self-reported measure of English proficiency (1=poor, 5=fluent). The final column provides a measure of each speaker’s intelligibility as measured by average number of words correctly perceived under four signal-to-noise ratios by a set of untrained listeners.

Seven female native speakers of American English were also recorded producing only the list of English words under the same conditions as the bilingual speakers. Productions from two of the female speakers were not included in the study due to problems these speakers had with completing the task accurately. The remaining five speakers were between the ages of 18-25 and reported no history of a speech or hearing disorder. These speakers received partial course credit for their participation.

Participants. Forty-two listeners participated in the German-training condition (21 in each training group) and 41 in the English-training condition (19 trained on group 1 talkers, 22 on group 2 talkers). All listeners were native speakers of American English attending Indiana University. In the German-training condition, 10 listeners were eliminated (did not reach criterion, N=5; did not complete the experiment, N=3; nonnative speaker of American English, N=1; lived in Germany, N=1), resulting in 32 usable listeners. Nine listeners were eliminated in the English-training condition (did not complete the experiment, N=4; nonnative speaker of American English, N=2; German-speaking parent, N=1; last participants to complete the experiment, N=2) yielding 32 usable listeners. None of the remaining 64 listeners reported any knowledge of German, had ever lived in Germany, or had any German-speaking friends or family members. All were between the ages of 18-25 and reported no history of speech or hearing impairments. Listeners were paid \$10/hour for their participation. In each training condition, half of the listeners were trained on Group 1 Talkers (“Group 1 Listeners”) and half on Group 2 Talkers (“Group 2 Listeners”).

The data from listeners who did not correctly identify at least 40% of the talkers in 3 (half) or more testing phases during training were excluded from analysis. This level of performance was selected to mirror the criterion used in Winters, Levi, and Pisoni (submitted). In addition, listeners were divided into “good learners” and “poor learners” following the criterion used in Nygaard and Pisoni (1998) who found that listeners who did not reach 70% accuracy in voice identification did not show the familiar talker advantage. In the German-training condition, 9/16 Group 1 Listeners and 7/16 Group 2 Listeners were classified as good learners. In the English-training condition, 8/16 Group 1 Listeners and 12/16 Group 2 Listeners were good learners.

Procedure. During the four days of the study, participants were seated in a quiet room at individual testing stations. All stimuli were presented to participants over Beyer Dynamic DT-100 headphones on PowerMac G4 computers running a customized SuperCard (version 4.1.1) stack. Participants were trained to identify one of two sets of five different bilingual voices by name in six training sessions spanning three days. Each training session consisted of seven distinct phases, summarized in Table 2. Each talker was associated with a common female name in both English and German and each name was presented in a different color in a unique position on the screen.

During each training session, listeners completed two training blocks followed by a testing block. Each training block began with two familiarization phases where listeners heard the same words from each of the five talkers. After familiarization, listeners completed a recognition task in which they heard five different tokens from each of the five talkers presented twice in random order. During recognition, listeners selected a talker by clicking an on-screen button next to the appropriate talker’s name and then received feedback by seeing the correct talker’s name and hearing the same stimulus token repeated again. After two training blocks, listeners completed a testing phase with no feedback. The testing phase consisted of 10 words produced once by each of the five speakers in random order. The only difference between the English and German training sessions was that the word used during familiarization B was the same as the last word used during familiarization A for the English trained listeners, but was a novel word for the German trained listeners. Each training session (consisting of two training blocks plus the test phase) lasted approximately 20 minutes. Participants completed two training sessions per day for three days. Participants were required to take a short (approximately five minute) break between consecutive sessions on each day of training.

Training Session			
	Phase	Stimuli	Task
Training Block I	Familiarization A	Same 5 words produced by each talker (500 ms ISI)	Listen and attend to talker-name pair
	Familiarization B	Same 1 word produced by each talker	Listen and attend to talker-name pair
	Recognition	5 different words produced by each talker, presented twice in random order	Identify speaker (feedback)
Training Block II	Familiarization A	same procedure as above	same procedure as above
	Familiarization B	same procedure as above	same procedure as above
	Recognition	same procedure as above	same procedure as above
	Test	10 different words produced by each talker, presented once in random order	Identify speaker (no feedback)

Table 2. Training procedure used for each session. Two training sessions were completed each day.

On the fourth day of the experiment, listeners completed a word recognition task in which they heard monosyllabic CVC English words and were asked to type what they heard. 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. Each stimulus was mixed with white noise which included a 200 ms linearly increasing ramp from 0 dB to the appropriate noise level at the beginning of the stimulus and a similar 200 ms decreasing ramp of the noise at the end. One quarter of the stimuli were presented at each SNR. No more than two days intervened between any of the four days of testing.

German-trained listeners heard all 360 English words during the word recognition task. One third of the stimuli were spoken by Group 1 talkers, one third by Group 2 talkers, and one third by the five native speakers of English. The English-trained listeners heard only 180 words during the word recognition task. These 180 words were randomly selected for each listener and they did not occur during any training sessions to avoid any lexical priming between the training and testing phases. For each listener, one third of these words were spoken by Group 1 Talkers, one third by Group 2 Talkers, and one third by the native English speakers.

Results

Training. An Analysis of Variance (ANOVA) was conducted on the response data from the test phases of the six training sessions. This ANOVA 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. In other words, listeners were able to learn the voices of the bilingual talkers across the three days of training. 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.

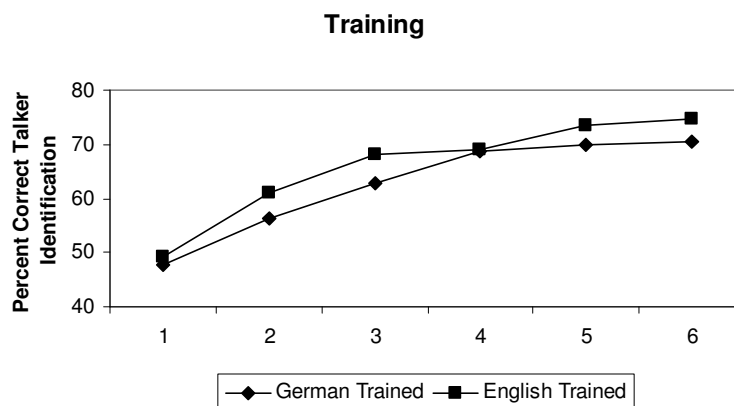


Figure 1. Talker identification accuracy during the six training sessions for both German-trained and English-trained listeners. Two training sessions were completed on each day of training.

Word Recognition. As previously mentioned, Nygaard and Pisoni (1998) found that “poor” learners did not exhibit a familiar talker advantage when performing a word recognition task similar to the one used in the current study. Using their criterion of 70% correct talker identification accuracy, listeners from both language training groups (English, German) and both talker training groups (trained on Group 1 talkers, trained on Group 2 talkers) were divided into “good learners” (those listeners who achieved 70% or greater on the last day of training) and “poor learners” (those listeners who did not reach 70% accuracy on the last day of training). Typed responses to the word recognition test were coded for whole word accuracy and for the number of correct phonemes per response (0-3). We first report the results for the German-trained listeners, followed by the English-trained listeners.

German-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. Figure 2 presents the results for whole words correct. For the good German-trained learners, only the main effect of SNR reached significance ($F(3,42) = 263.6, p < .001$), indicating that listeners performed better under more favorable SNRs. 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 also 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 (45.1% vs. 42.7% words correct). This difference in average intelligibility for the poor learners likely reflects the inherent intelligibility differences in the two groups of talkers (see Table 1).

Similar results were obtained for the number of phonemes correctly identified during word recognition (Figure 3). 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 ($F(1,14) = 6.104, p = 0.027$). As with the whole word correct data, poor German-trained learners perceived more phonemes correctly for Group 2 talkers than Group 1 talkers (70.4% vs. 68.8% phonemes correct).

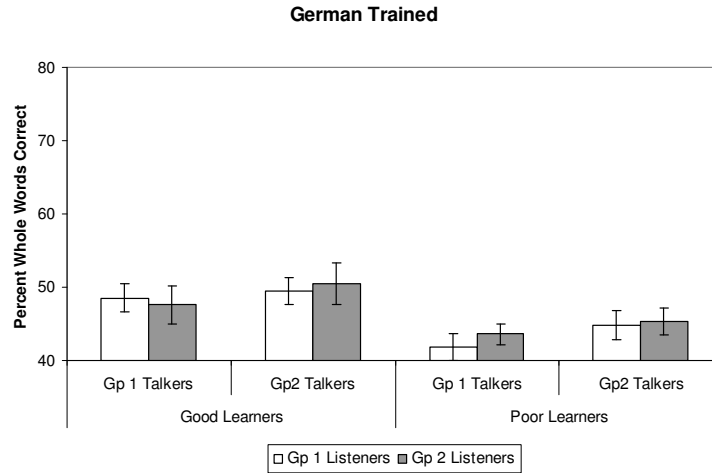


Figure 2. Percent whole words correct for German-trained listeners.

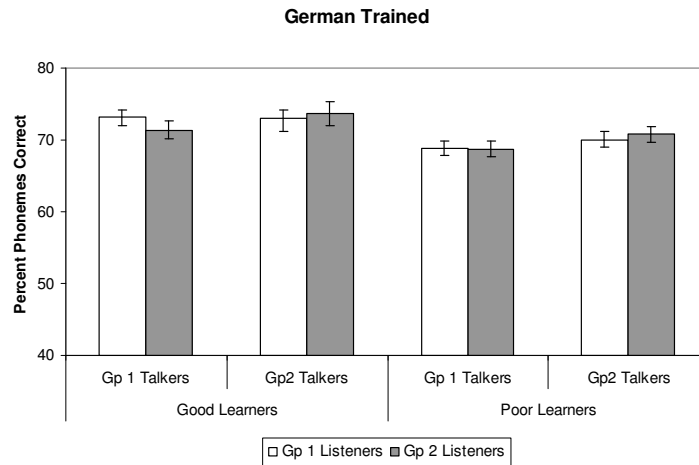


Figure 3. Percent phonemes correct for German-trained listeners.

English-trained Listeners. The pattern of results for the English-trained listeners differs from of the results obtained for the German-trained listeners. As with the German-trained listeners, separate ANOVAs for good and poor learners were conducted on the whole word correct data with Talker Group and SNR as within-subjects factors and Listener Group as a between-subjects factor. For the good English-trained learners, a main effect of SNR was found ($F(3,54) = 2.14.8, p < .001$). In addition to this main effect, the talker group by listener group by SNR interaction also reached significance ($F(3,54) = 2.918, p = .041$) and the talker group by listener group interaction approached significance ($F(1,18) = 3.632, p = .071$). 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 4 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 results at each SNR, driven mostly by a large benefit of 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 ($F(3,30) = 339.7, p < .001$).

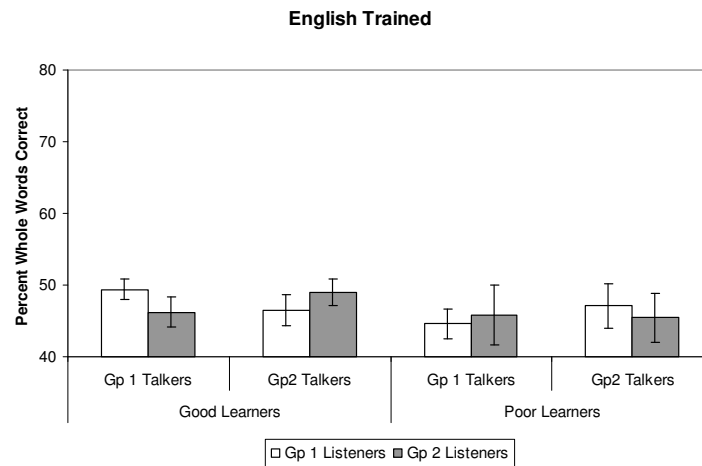


Figure 4. Percent whole words correct for English-trained listeners.

Similar results were found for the number of phonemes correctly identified during word recognition (Figure 5). For the good English-trained learners, a main effect of SNR was found ($F(3,54) = 8.654, p < .001$). In addition, the talker group by SNR interaction ($F(3,54) = 3.121, p = .032$) and the talker group by listener group interaction ($F(1,18) = 8.674, p = 0.008$) were significant. 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. 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. For the poor learners, the main effect of SNR reached significance ($F(3,30) = 5.321, p < .001$), as did the talker group by listener group by SNR interaction ($F(3,30) = 3.597, p = .025$). 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.

Correlational Data. Converging evidence for the differences between English-trained and German-trained listeners and additional support for separating listeners into good and poor learners was obtained from correlations carried out between the degree of talker familiarity and performance on the word recognition task. Bivariate correlations were conducted between each listener's talker identification score (percent of talkers correctly identified on the last day of training) – a measure which indicates degree of familiarity with the talkers – and the observed gain in speech intelligibility, which was computed as the difference between trained talkers and untrained talkers in the word recognition task. Separate correlations were run for the whole word correct data and the phonemes correct data. For German-trained listeners, no significant correlations were found for either whole words correct or phonemes correct. In contrast, significant correlations were found for both whole words ($r = .353, p = .041$) and phonemes ($r = .466, p = .006$) correct for the English-trained listeners. The correlations found between degree of talker familiarity and intelligibility gain for the English-trained listeners indicates that listeners who were more familiar with the talkers – as measured by higher identification scores –

exhibited greater gains in speech intelligibility for familiar talkers. Thus, the better listeners are at learning a talker's voice, the better they are at recognizing spoken words from that talker. This positive correlation was only found with listeners who were trained with English stimuli. German-trained listeners did not exhibit this correlation, corroborating data in the previous sections which showed no familiar talker advantage for these listeners. Taken together, all of these results indicate that the familiar talker advantage is only found for English-trained listeners and that they show a relationship between degree of talker-familiarity and intelligibility gain.

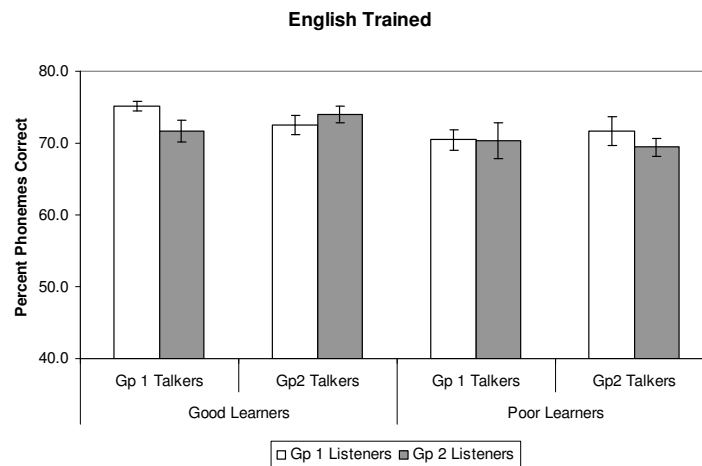


Figure 5. Percent phonemes correct for English-trained listeners.

General Discussion

The present study demonstrated that the familiar talker advantage only occurs if the same language is used during perceptual learning and word recognition, in particular where the talker learning task incorporates a linguistic component; English-speaking listeners learning voices in English do not avoid or suppress word recognition during the talker learning phase and therefore also perceive the linguistic content of the speech. These listeners showed the expected familiar talker advantage by correctly perceiving more words and more phonemes for trained talkers than for untrained talkers. In contrast, listeners who were trained on the same bilingual talkers, but using German stimuli, did not show any benefit of talker familiarity during the linguistic task. Correlational data corroborated these findings, showing that greater familiarity with a set of talkers is associated with a larger familiar talker advantage for English-trained listeners but not for German-trained listeners. Thus, it appears that increased performance in a linguistic task is contingent upon being familiar with a talker's linguistically-relevant productions in that language and upon learning the link between talkers and their linguistic characteristics.

We attribute these findings to the overlap and type of indexical information that are available in these two tasks. Listeners trained on English stimuli acquired both language-independent and English-specific indexical properties about the individual talkers, both of which are also present in the English stimuli used during word recognition. In contrast, listeners trained on German stimuli learned only language-independent indexical properties – and some German-specific indexical properties – but crucially not English-specific indexical properties. When these listeners performed a word recognition task with English stimuli, the only talker information that they had stored was language-independent

information, which contains no linguistic information that could facilitate English word recognition. We conclude that knowledge of language-specific indexical properties is necessary to generate a familiar talker advantage because listeners must know linguistically relevant talker information (i.e., language-specific indexical information) to display gains in linguistic processing.

The absence of a familiar talker advantage for the German-trained listeners cannot be explained by arguing that the talkers are perceived as unfamiliar in English. The existence of language-independent indexical properties which make talkers identifiable across languages has been established in work on cross-language talker identification and discrimination (Winters et al., submitted). Winters et al. found that listeners trained to identify talkers in one language were able to identify them in another language at levels well above chance. Furthermore, Winters et al. showed that untrained listeners can reliably discriminate talkers when speaking different languages. Thus, the lack of a familiar talker advantage by German-trained listeners is not due to the talkers sounding unfamiliar in English, because talkers can be identified from English stimuli by German-trained listeners with no decrease in performance. Rather, the absence of a familiar talker advantage must be attributed to the lack of learned English-specific indexical properties.

In the remainder of this section we explore why knowledge of language-specific indexical properties is necessary to generate the familiar talker advantage by examining data from perceptual learning studies, bilingual speech production, and cross-modal talker familiarity. We then briefly introduce two theories that have been used to account for the effects of talker variability on linguistic processing and discuss how these same perceptual mechanisms also explain the benefit of familiar voices on linguistic processing by English-trained listeners but not German-trained listeners.

Source of the Familiar Talker Advantage

When listeners learn a talker's voice from English stimuli, they also perceive the linguistic content of the utterance and thus acquire valuable information about how a talker articulates specific linguistic contrasts. Data from several perceptual learning studies show that listeners encode and retain talker-specific information about speech production and that this knowledge modifies how listeners perceive linguistic contrasts (Allen & Miller, 2004, Eisner & McQueen, 2005; Kraljic & Samuel, 2005). In one study, Allen and Miller (2004) trained listeners on the voices of two talkers, one with long voice onset times (VOTs) and one with short VOTs. During the test phase, listeners generalized talker-specific VOT differences to novel words, indicating that listeners' sensitivity to subphonemic, acoustic-phonetic differences was retained in memory and used in language processing tasks in a talker-contingent manner. Similarly, Eisner and McQueen (2005) and Kraljic and Samuel (2005) reported talker-specific subphonemic attunement in a fricative perception task. Eisner and McQueen trained listeners with an ambiguous fricative in either an [f]- or [s]-biasing lexical context and then asked them to categorize stimuli along an f/s continuum. Listeners' category boundaries were only shifted for stimuli produced in the same voice as the training stimuli. In another study using ambiguous [s] and [ʃ] stimuli, Kraljic and Samuel showed that perceptual learning of talker-specific characteristics is retained up to at least 25 minutes. All of these earlier studies provide clear and consistent evidence that listeners encode and retain talker-specific, linguistically-relevant production information for different talkers. Furthermore, this talker-contingent knowledge alters how listeners perceive the speech of familiar talkers, showing that listeners' category boundaries are talker-dependent.

Further evidence for why language-specific indexical information is necessary to elicit the familiar talker advantage comes from production studies of bilingual speakers. Languages which contain the "same" phonological contrast do not necessarily use the same cues or the same category boundary to

differentiate segments. For example, VOT values for stop consonants in Canadian French are shorter than for Canadian English (voiceless: 37 ms vs. 88 ms; voiced: -99 ms vs. 20 ms) (MacLeod & Stoel-Gammon, 2005). This difference in monolingual production is largely maintained in the speech of bilinguals, who use language-appropriate VOTs when speaking the different languages (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; MacLeod & Stoel-Gammon, 2005). From these findings it is clear that knowledge of how a talker articulates a linguistic contrast in one language will not necessarily help to perceive a linguistic contrast in another language because the location of a talker's category boundary, as well as the types of cues used to distinguish different segments, are language-dependent. Thus, for the German-trained listeners in our study, learned German-specific indexical properties will not help them perceive a familiar talker's speech in English.

Finally, evidence that listeners need exposure to language-specific (i.e., English-specific) indexical properties to exhibit a familiar talker advantage comes from a recent study of cross-modal talker facilitation. Rosenblum, Miller, and Sanchez (2007) had participants first transcribe sentences from visual-only stimuli and then transcribe novel sentences from auditory-only stimuli produced by either the same or different talker. Although participants were not explicitly directed to attend to talker characteristics during the initial visual-only transcription task, participants nonetheless perceived more words correctly when the same talker was used in both the visual-only and auditory-only tasks. Crucially, familiarization and testing were conducted in the same language, thus providing participants with language-specific indexical information. Although some learned indexical information in Rosenblum et al.'s study was non-acoustic, the gestural articulations of contrasts were specific to English. This knowledge of English-specific articulations facilitated speech perception of a familiar talker across modalities.

Taken together, these recent studies present converging evidence that listeners must have prior knowledge of language-specific indexical information for the familiar talker advantage to be observed. In addition, the perceptual learning studies and bilingual production studies suggest that this knowledge is necessary to enhance linguistic performance when listeners are asked to recognize words mixed in noise. When listeners are familiar with how a talker produces linguistic contrasts, they make fewer errors in (linguistic) perception. Furthermore, when listeners are only familiar with a talker's production in a different language (e.g., German), there is no facilitation of linguistic processing, because listeners lack the relevant language-specific knowledge about the talker. We now consider two accounts for why familiar voices are processed more quickly and accurately than unfamiliar talkers.

Relationship between Indexical and Linguistic Processing

Two accounts have been proposed to explain the link between indexical and linguistic processing and in particular to explain the adverse effects of talker variability on linguistic processing. Exemplar models (e.g., Goldinger, 1998, Johnson, 1997) assume that the processing cost associated with talker variability exists because both linguistic and indexical information are transmitted through the same stream of information (i.e., the acoustic waveform) and because both of these types of information are simultaneously retained in an exemplar stored in memory. In contrast, normalization theories (e.g., Nusbaum & Magnuson, 1997, Magnuson & Nusbaum, 2007) assume that each change in the talker dimension requires listeners to continually adjust and readjust their perceptual system to map a different talker's utterances onto the correct linguistic target, thus slowing perception and increasing the likelihood of errors.

Although these theories have been primarily used to explain the effects of talker variability on linguistic processing, they can also readily account for the facilitation due to talker familiarity and why

this facilitation is only observed for the English-trained listeners in our study. In exemplar theories, familiar talkers are represented by more exemplars stored in memory. The more exemplars that exist for a particular talker, the more likely it is that an incoming stimulus will match these stored exemplars along various linguistic dimensions facilitating linguistic processing. Thus, listeners trained on English stimuli exhibit fewer errors when listening to familiar talkers. In contrast, listeners trained on German stimuli only store German exemplars and examples of German linguistic contrasts. For these listeners, an incoming English stimulus from a familiar talker is not inherently more similar to existing English exemplars than a stimulus from an unfamiliar talker because no English exemplars exist for either talker. Therefore, German-trained listeners should not exhibit a familiar talker advantage.

In theories of talker normalization, listeners actively adjust their perceptual system to talker differences and learn to process the linguistic content of an utterance in a talker-contingent manner. For familiar talkers in English, the path between a speech stimulus and the linguistic abstractions is well-paved because listeners have abundant experience interpreting a familiar talker's speech from previous experience. If a listener is familiarized with talker in a different language, then the process of recognizing the talker and mapping his/her utterance onto a linguistic representation is never completed. Because German-trained listeners cannot create the mapping between talker and linguistic content, linguistic processing of "familiar" talkers is no different from unfamiliar talkers; both require a new processing strategy. Whichever theory is ultimately shown to best explain the interaction between indexical and linguistic processing, the important point here is that the same mechanisms that account for talker variability effects can also be used to account for the familiar talker advantage observed with English-trained listeners and the lack of this advantage for German-trained listeners.

Conclusions

Using a cross-language voice learning paradigm, we sought to explore the underlying causes for the familiar talker advantage. To this end, we manipulated the type of information available to listeners by familiarizing one group of listeners with all potentially relevant talker characteristics (language-independent and English-specific indexical properties) and a second group with a limited amount of talker characteristics (language-independent indexical properties). The group of listeners trained on English stimuli showed the expected familiar talker advantage during word recognition, whereas the group trained with German stimuli did not. The results of this study provide additional evidence that linguistic processing is performed in a "talker-contingent" manner and that listeners must have knowledge of talker-specific linguistic information to facilitate linguistic processing in a word recognition task. The absence of a familiar talker advantage for the German-trained listeners demonstrates that the familiar talker advantage is not due to knowing a voice *per se* or to being able to identify or discriminate different talkers, but rather to knowing how a voice (talker) produces linguistically significant contrasts in the language. Thus, to show an advantage in linguistic processing, a listener must acquire linguistic knowledge about the talker's speech.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine Publishing Company.
- Allen, J.S. & Miller, J.L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 115, 3171-3183.
- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

- Caramazza, A., Yeni-Komshian, G.H., Zurif, E.B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 54, 421-428.
- Eisner, F. & McQueen, J.M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67, 224-238.
- Goggin, J.P., Thompson, C. P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, 19, 448-458.
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.
- 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.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J.W. Mullennix (Eds.) *Talker variability in speech processing* (pp. 145-164). San Diego: Academic Press.
- Köster, O., & Schiller, N.O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, 4, 18-28.
- Kraljic, T. & Samuel, A.G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.
- Levi, S.V., Winters, S.J., & Pisoni, D.B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, 121, 2327-2338.
- MacLeod, A.A.N. & Stoel-Gammon, C. (2005). Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders*, 3, 118-127.
- 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.
- Mullennix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.
- 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.
- Nusbaum, H. & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J.W. Mullennix (Eds.) *Talker variability in speech processing* (pp. 109-132). San Diego: Academic Press.
- Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355-376.
- Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- 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.
- Rosenblum L.D., Miller R.M., & Sanchez, K. (2007). Lip-read me now, hear me better later. *Psychological Science*, 18, 392-396.
- 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.
- Schiller, N.O., & Köster, O. (1996) Evaluation of a foreign speaker in forensic phonetic: a report. *Forensic Linguistics*, 3, 176-185.
- Schlichting, F. & Sullivan, K.P.H. (1997). The imitated voice – a problem for voice line ups? *Forensic Linguistics*, 4, 148-165.

- Sullivan, K.P.H., & Schlichting, F. (2000). Speaker discrimination in a foreign language: first language environment, second language learners. *Forensic Linguistics*, 7, 95-111.
- Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, 1, 121-131.
- Winters, S.J., Levi, S.V., & Pisoni, D.B. (submitted). Identification and discrimination of talkers across languages. *Journal of the Acoustical Society of America*.

