L2 PERCEPTION OF ENGLISH FRICATIVES IN CLEAR AND CONVERSATIONAL SPEECH: THE ROLE OF PHONEMIC, PHONETIC, AND ACOUSTIC FACTORS

Bariş Kabak, Kazumi Maniwa

Department of Linguistics, University of Konstanz, Germany

Baris.Kabak@uni-konstanz.de, Kazumi.Maniwa@uni-konstanz.de

ABSTRACT

This study investigated perception by non-native listeners of English fricatives produced in clear and conversational speaking styles. We measured babble thresholds for fricative voicing and place of articulation contrasts by Standard German and Swabian German and native American English speakers. Overall, Swabian German speakers performed worse than both native English and Standard German speakers, and Standard German speakers worse than native English speakers. German speakers in general had more difficulty with non-sibilant distinctions, and Swabian speakers also had difficulty with sibilant voicing distinctions. A robust clear speech benefit was observed across groups and contrasts. Overall, the results indicate that difficulty in perceiving foreign-language contrasts stems from the interaction of phonological, phonetic, and psychophysical issues.

1. INTRODUCTION

Adults' perception of non-native speech contrasts is strongly influenced by the phonological systems of their native languages [e.g.1, 2]. The phonemic status of a foreign contrast in a speaker's native language is clearly important in influencing perception, but it alone does not explain why some novel contrasts appear to present greater perceptual difficulty than others, or why speakers of different language backgrounds have differing difficulty with the same novel contrast. One factor that may play a role is the degree of similarity between non-native sounds and sounds in speakers' native inventories [e.g.3, 4]. It has also been suggested that experience with foreign phonetic categories as allophones or as free variants of native phoneme categories is important [e.g.5, 6].

A very different possibility is that difficulty with new contrasts is determined by the psychophysical salience of the acoustic cues that serve to differentiate the sounds [e.g.7, 8]. Burnham [8] proposed a distinction between "robust" contrasts, which are common across languages, acquired early, maintained by infants in the absence of phonemic experience, and easily learned by adults; and "fragile" contrasts, which are less salient in these respects.

Both interactions with speakers' native inventories and issues of acoustic/perceptual salience probably contribute to difficulty in the perception of non-native sounds. However, cross-language research has not thoroughly explored the interrelations among these factors in L2 perception but has primarily focused on L1-L2 phonological interactions. This study seeks to address the relative importance of phonemic, phonetic, and acoustic factors by observing the perception of English fricatives by standard German and Swabian German listeners.

In Standard German, the voicing distinction for sibilant fricatives, a "robust" contrast, is very restricted. Although a few minimal pairs can be found (e.g., *rei[s]en* vs. *rei[z]en*), [s] and [z] mostly occur in complementary distribution: [z] appears syllable-initially before a vowel whereas [s] occurs syllable-finally due to coda devoicing. Voiced palato-alveolar /ʒ/, on the other hand, can be found only in loan words, and is considered to be a "peripheral" phoneme since it occurs neither in contrastive distribution nor in free variation with [ʃ] [9]. The Swabian German dialect does not have voiced sibilants even allophonically, so the voicing distinction is completely absent for sibilants. Neither Standard nor Swabian German has interdental nonsibilants, which participate in "fragile" contrasts with other sounds in English and German (in particular with bilabial fricatives, /f/ and /v/, which are contrastive in both varieties). Thus, a comparison of Standard and Swabian German speakers' perception of sibilant voicing and non-sibilant place distinctions with (i) perception of these distinctions by native English speakers and (ii) perception of robust and fragile contrasts in the German speakers' inventory (i.e. sibilant place and non-sibilant voicing distinctions) provides an invaluable opportunity to assess the relative importance of phonemic, phonetic, and acoustic factors in cross-language speech perception.

A secondary purpose of this study was to test whether naturally produced "clear speech" provides effective intelligibility enhancement for non-native listeners. Clear speech is more intelligible than conversational speech for a variety of (native) listener populations under various listening conditions [e.g.10, 11]. However, clear speech perception data for non-native listeners is limited, and benefits may be smaller. Bradlow and Bent [12] interpreted a relatively small "clear speech effect" for non-native listeners as a consequence of clear speech being native-listener oriented, and therefore mostly beneficial to listeners with extensive experience with the structure of the target language. Investigating whether the perception of non-native contrasts can be enhanced by clear speech may help to clarify both the nature of clear speech and the sources of difficulty in perceiving non-native contrasts. In particular, robust clear speech effects for non-native listeners would suggest both that clear speech serves to enhance the acoustic salience of contrasts and that acoustic salience is important in the perception of non-native contrasts. It is unlikely that non-native listeners would show large clear speech benefits if clear speech consists of primarily native listener-oriented alternations, or if they are relying primarily on the relationships between novel sounds and contrasts and their native inventory, since these relationships would not change predictably in clear speech.

2. EXPERIMENT

2.1 Participants

14 normal-hearing Standard-German listeners (8F, 6M) aged 18 to 30 (M=24.79 yrs old) and 14 normal-hearing Swabian-dialect listeners (10F, 4M) 19 to 27 (M=22.29) were recruited from the University of Konstanz community. Participants had studied English for a minimum of 7 years. Additional information about the group of non-native subjects is provided in Table 1. 14 normal-hearing native speakers of American English (8F, 6M) 19 to 32 (M=24.56) were recruited from the University of California, Berkeley.

	Standard Mean	Range	Swabian Mean	Range
Age	**24.79**	**18-30**	**22.29**	**19-27**
Length of English Study	**10.36**	**7-14.5**	**9.35**	**7-13**
Length of time in an English speaking community	**5.71 months**	**0-24**	**3 months**	**0-14**
Reading	**1.57**	**1-3**	**2.21**	**1-4**
Writing	**2.29**	**1-3**	**2.57**	**2-3**
Speaking	**2**	**1-3**	**2.86**	**1-4**
Comprehension	**1.64**	**1-3**	**2.21**	**1-3**

Table 1: *General information about nonnative subjects: Reading, Writing, Speaking, and Comprehension skills were self-rated from 1 (very poor) to 5 (very good)*

2.2 Stimuli

Fricative contrast perception was assessed using a database of 8800 VCV ([a]-fricative-[a]) stimuli produced by 20 speakers (10 F, 10 M) as part of a previous acoustic study of clear fricatives [13]. Briefly, conversational and clear tokens were elicited using an interactive program that ostensibly attempted to recognize fricatives produced by speakers. The program made frequent, systematic errors involving voicing and place alternations, after which a speaker repeated a sound more clearly, as if trying to disambiguate the production for a listener. Stimuli were presented in a background of 12-talker (6M, 6F) babble. Target VCV were centered temporally in randomly selected babble segments exceeding the target duration by 600 ms. There were 5-ms and 100-ms linear on-off ramps for the target stimulus and the noise, respectively.

2.3 Procedure and data analysis

The perception test employed a two-alternative forced-choice identification task. The 8 fricatives were divided into 8 minimal pairs, depending on place of articulation and voicing: /f/-/θ/, /v/-/ð/, /s/-/ʃ/, /z/-/ʒ/, /f/-/v/, /θ/-/ð/, /s/-/z/, and /ʃ/-/ʒ/. Each pair was tested separately for clear and conversational styles, for a total of 16 sub-tests. Contrasts /s/-/ʃ/ ('robust' contrast) and /f/-/v/ ('fragile' contrast) are phonemic for all three listener groups and as such may be thought of as 'control pairs' in which differences in SNR among groups are not expected. The goal of each test was to determine the signal-to-noise ratio (SNR) threshold at which a distinction could be made with 75% accuracy. On each trial, a test VCV and a segment of babble noise were randomly selected. The two waveforms were scaled based on the selected SNR and a constant target stimulus level, combined additively, and presented binaurally to subjects, who identified the fricative from a minimal pair using a mouse to click one of 2 letter combinations on the computer screen. Test order was randomized across subjects in a single 1-hour session. Before each test, listeners were oriented to the spelling of response alternatives using a 10-trial initiation test at a high SNR (+10dB) with feedback. Within tests, two randomly interleaved 40-trial adaptive tracks were initiated at +3dB and -3dB. SNR values for each track were selected using a Bayesian adaptive algorithm [14]. The final threshold estimate was taken as the average of the two tracks' SNR values on the final trial.

Effects of speaking style, contrast, and native language/dialect were tested using a repeated measures ANOVA with two within-subject factors (Style; 2 levels, Pair; 8 levels), listener group as a between-subject factor (Standard, Swabian, American) and threshold (dB SNR) as the dependent variable. Pairwise comparisons for significant within-subject factors used Bonferroni corrected 95% confidence intervals.

3. RESULTS AND DISCUSSION

Examination of individual results revealed that all subjects performed with approximately 75% accuracy during the final 5 trials of a block, indicating that the constant 80-trial block length was sufficient to estimate thresholds. Figure 1 shows mean SNR thresholds (dB) as a function of fricative pair and speaking style for the three subject groups. Style effect ($p<.0001$) indicated lower thresholds for clear speech than for conversational speech. The pair effect was also significant ($p<.0001$); across speaking style and group, thresholds were lowest for the voiceless sibilant place of articulation distinction /s/-/ʃ/, followed by /z/-/ʒ/, /s/-/z/, and /ʃ/-/ʒ/. The interdental place distinctions had the highest thresholds in both speaking styles and for all groups. This suggests that that poor identification of interdental consonants is influenced by acoustic, rather than language-specific, issues. The Style × Pair interaction ($p<.0001$) showed that sibilant pairs improved in clear speech to a greater extent than nonsibilant pairs, although intelligibility significantly improved for all pairs across groups. An effect of Group ($p<.0001$) showed that overall, Standard German speakers performed worse than native, and Swabian speakers worse than both other groups. A Pair × Group interaction ($p<.0001$) revealed that effect resulted from better performance by native speakers on distinctions involving interdentals, and worse performance by Swabian speakers for sibilant voicing contrasts. The differences in interdental distinctions are consistent with a phonemic influence on non-native speech perception. In addition, the difference in performance between Swabian and Standard German listeners for sibilant voicing distinctions provided evidence for phonetic/ allophonic influence.

There was no Style × Group interaction; on average, all listener groups benefited similarly from clear speech. This differs somewhat from previous findings [12] which suggest that the clear speech effect is greatly reduced for non-native listeners. It might be that the small clear speech effects previously observed in non-native listeners were mostly due to that the use of meaningful sentences in that study, including context that non-native listeners would be less able to take advantage of. It also emphasizes the role of general acoustic enhancement in clear speech, and of the effects of acoustic salience in perception of non-native contrasts.

While clear speech improved perception for all pairs and all groups, German listeners did not approach native-like perception for non-sibilants, nor Swabian speakers for sibilant voicing distinctions, indicating a strong influence of their L1 phonological and phonetic systems. In fact, a separate repated measures of ANOVA within subject groups showed that the clear speech intelligibility benefit was greater for sibilant voicing distinctions than for interdental voicing and place-of-articulation distinctions for Swabian German listeners ($p<.01$). This supports the influence of acoustic salience in cross language speech perception: "robust" contrasts such as /s/-/z/ are more easily improved than "fragile" contrasts such as /f/-/θ/.

To examine the influence of familiarity and ability in English on speech perception, correlation coefficients (Pearson's r) were assessed between clear, conversational, and clear-minus-conversational differences and demographic variables (i.e. length of English study, length of time in an English speaking community, and self-rated reading, writing, and comprehension for the non-native listeners).

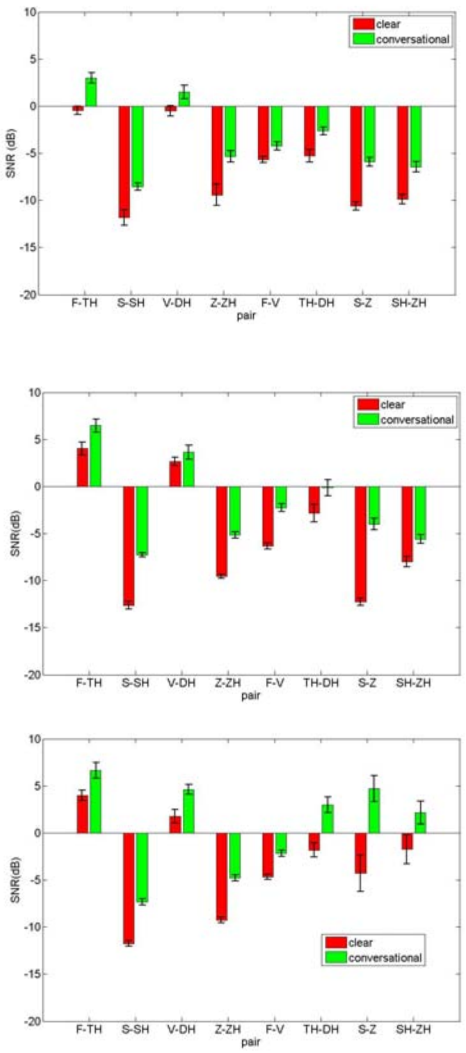

Figure 1: *Speech-to-noise (SNR) thresholds (dB) as a function of style and fricative pair for native American English (top), Standard German (middle), and Swabian German (bottom) listeners*

There was in general a complete lack of correlation between any of these variables. This differs somewhat from previous results based on more experienced L2 speakers [e.g. 4], but is consistent with the variability seen in studies of listeners with similarly limited experience with the target language [e.g. 12]. Further research will be required to determine whether some of these factors contribute to the effects reported here.

In sum, the different performances obtained in the non-native groups reflect phonological encoding and exposure to the phonetic context of voicing in different L1s, although there are also general perceptual issues due to acoustic salience for all groups. Thus, it is interactions between these factors (and probably others), and not any one factor acting alone that determines how listeners perceive non-native contrasts.

4. REFERENCES

[1] Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In: Archibald, J. (ed.), *Second Language Acquisition and Linguistic Theory*. Malden, MA: Blackwell, 4-63

[2] Werker, J. F., Gilbert, J. H., Humphrey, K., Tees, R. C. 1981. Developmental aspects of cross-language speech perception. *Child Dev.*52, 349-353

[3] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 171-206

[4] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 233-273

[5] Ingram, J. C. L., Park, S-G. 1998. Language, context, and speaker effects in the identification and discrmination of English /r/ and /l/ by Japanese and Korean listeners, *J. Acoust. Soc. Am.* 103, 1161-1174

[6] Harnsberger, J. D. 2000. A cross-language study of the identification of non-native nasal consonants varying in place of articulation, *J. Acoust. Soc. Am.* 108, 764-783

[7] Polka, L. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *J. Acoust. Soc. Am.* 89, 2961-2976

[8] Burnham, D. K. 1986. Developmental loss of speech perception: Exposure to and experience with a first language. *Appl. Psycholinguist.* 7, 207-239

[9] Benware, W. 1986. *Phonetics and Phonology of Modern German: An introduction*. Washington, D. C.: Georgetown University Press

[10] Picheny, M. A., Durlach, N. I., Braida, L. D. 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech*. J. Speech Hear. Res.* 28, 96-103

[11] Helfer, K. 1998. Auditory and auditory-visual recognition of clear and conversational speech by older adults. *J. Am. Acad. Audiol.* 9, 234-242

[12] Bradlow, A. R., Bent, T. 2002. The clear speech effect for non-native listeners. *J. Accoust. Soc. Am.* 112, 272-284

[13] Maniwa, K. 2007. Acoustic and perceptual properties of clearly produced fricatives, Unpublished Doctoral Dissertation.

[14] King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., Supowit, A. 1994. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res.* 34, 885-912