TRAVAUX DU CERCLE LINGUISTIQUE DE COPENHAGUE

VOL. XXXI

Nina Grønnum and Jørgen Rischel (eds.)

TO HONOUR ELI FISCHER-JØRGENSEN

C.A. REITZEL
COPENHAGEN 2001

Hearing the Polish sibilants [s ś š]: Phonetic and auditory judgments

Leigh Lisker

1. Introduction

While fricative consonants are usually lumped together with stops as »true« consonants, they resemble both the stops and other segment types in their perceptual properties. As in the case of the stops, some place information is provided by transitional shifts in formant patterns adjacent to the interval of constriction, but unlike the stops, much place information is also provided by acoustic properties of the sound emitted during the constriction. Moreover, at least some of the fricatives seem to be continuously variable in "color," more like vowels than stops.[1] Certain experiments reported by K. S. Harris (1958) indicate that for English [sʃfθ] the contributions of noise intervals and transitions to perception are variable, the noises being of overwhelming importance for the "strident" consonants [s] and [ʃ], and transitions playing a greater role in the case of the less noisy [f] and [θ]. The different perceptual weightings of noise and transition seem readily explained by the intensity relations between the two. Of course, a spectrographic look at the syllables /sɑ/ and /šɑ/ shows them as differing markedly in transitions, just as the /f/ and /θ/ noises show spectral differences. One motive for initiating the study reported here was to see whether, for English-speaking listeners, the relation between noise and transition cues to the English fricatives holds more generally for this class of phonetic segments, even those drawn from a language such as Polish, which is rather richer in such segments than is English, particularly in regard to sibilants. Where English has [s] and [š], Polish has three: [s], [š] and [ś] (described as post-dental, palato-alveolar and alveolo-palatal, or more tersely as simply dental, alveolar and palatal (Jassem 1964). The first two are said to be very like English [s] and [š] (Shenker 1973), al-

though the third also sounds pretty much like English [š] to English speakers. But before addressing the question of how and on what basis English-speaking listeners might categorize the Polish sounds, let us look at three of their visible transforms.

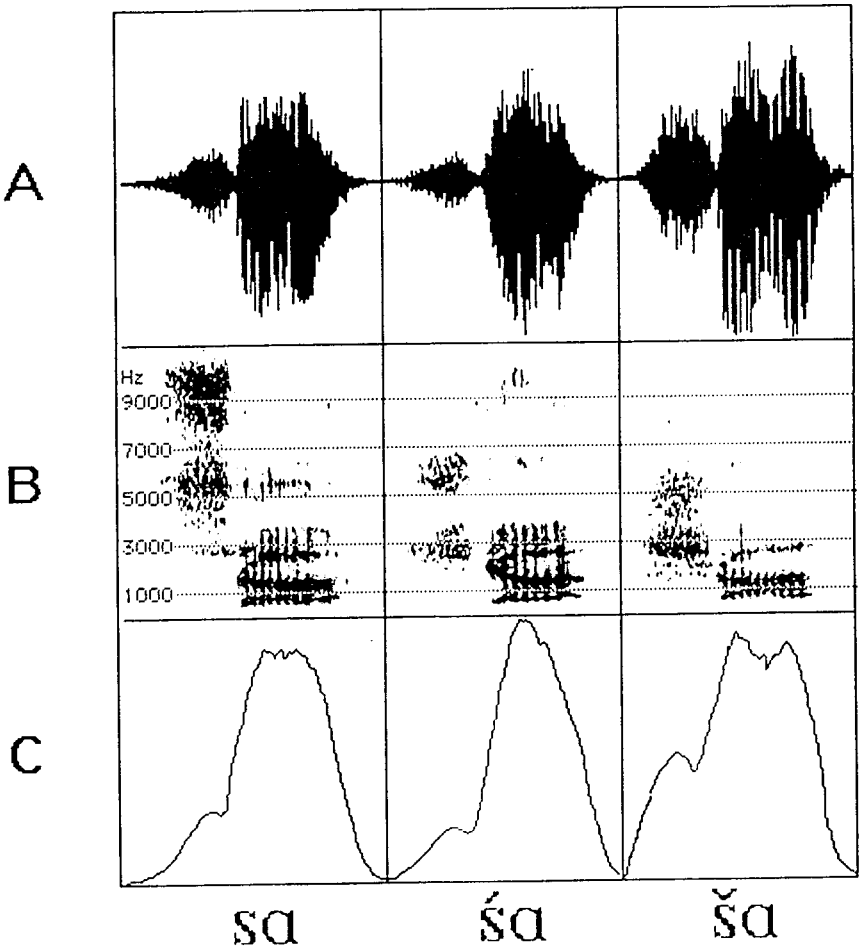

Figure 1. Acoustic displays of representative tokens of Polish syllables [sa], [śa] and [ša]: (A) waveforms, (B) spectrograms, and (C) amplitude profiles.

2. Some simple acoustic observations

In Figure 1 we see pressure waveforms, spectrograms, and amplitude contours of representative tokens of the three nonsense syllables [sɑ], [śɑ] and [šɑ][2] pronounced by an adult male speaker of standard Warsaw Polish. Four tokens of each syllable type were produced in isolation in randomized order and recorded in a suitably shielded sound booth. These syllables were presented for identification to two native speakers of Polish, whose responses showed that each syllable was identified exactly as the speaker had intended. A cursory examination of the spectrograms suggests at least two visual (acoustic) classifications of the three syllables,– one on the basis of frequency characteristics of the noise spectra, the other based on the shape of the second formant transition. From the representative averaged spectra of the three noises (Figure 2), it appears that [s] has a major concentration of energy in the

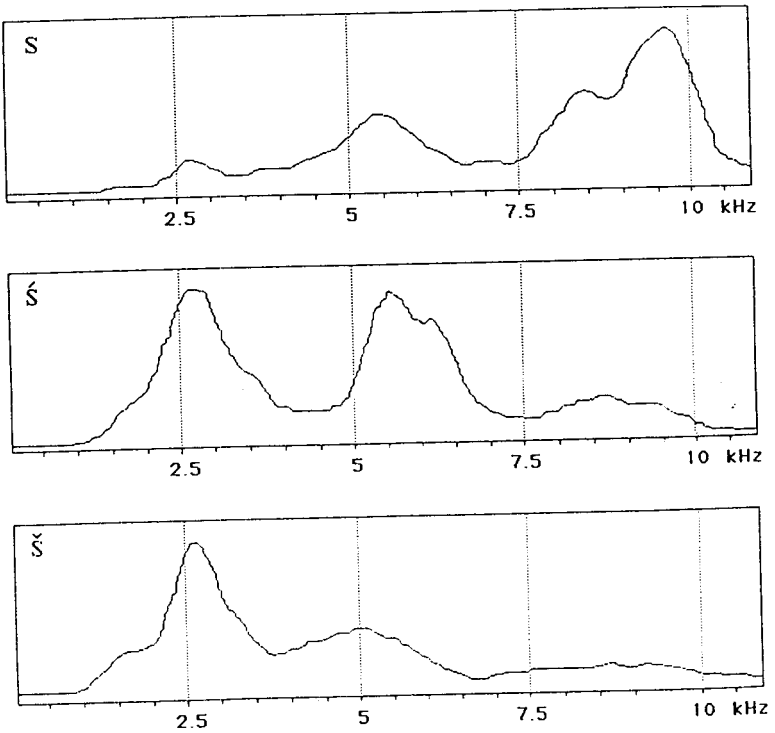

Figure 2. Spectra of representative [s], [ś] and [š] fricative noises averaged over their entire durations.

neighborhood of 9 kHz., with little below 5 kHz, while in both [š] and [ś] energy peaks are found at about 6 kHz and lower. Spectrographically, the [śɑ] syllable shows an F2 vocalic transition that begins at a considerably higher frequency than do those in either [sɑ] or [šɑ]. F2 transitions of [sɑ] and [šɑ] look remarkably alike, despite the recognized fact that the two categories differ in place of articulation, and that the second formant transition is often said to be a most important cue to this sort of difference.[3] (There are other differences to be seen. For one, both the spectrograms and the waveforms suggest that the three noises might have characteristically rather different amplitudes relative to the following vowels: [s] < [ś] < [š] (but see Strevens 1960). A closer look at the transitional patterns suggests the possibility of perceptually significant differences in the other formants, the third and perhaps the first, fourth, and even higher ones.)

3. Phonetic labeling tests

3.1 Procedures

Because classification by eye is no substitute for one by ear, several kinds of perceptual tests were carried out, the stimuli being presented at a comfortable level by loudspeaker in a sound-proof studio. First of all, an informal pilot test of five phonetically naive American English speakers fully bore out the expectation that Polish [s] would be equated with English [s], and that both the Polish alveolar (or palato-alveolar) [š] and palatal (or alveolo-palatal) [ś] would be identified largely with English [š].[4] More serious testing was thereafter confined to a dozen English speakers with some degree of linguistic training, but no previous experience of Polish or its alveolar-palatal shibilant contrast. After a brief initial exposure to a randomized sequence composed of one token of each of the three syllable types, where listeners were provided with feedback as to the correctness of their judgments, they were then presented with randomly ordered sets of stimuli derived from four tokens of each type, with an interstimulus interval of four seconds between each stimulus (each heard twice in immediate succession, and with eight exposures of each token). Three tests were carried out, with different orders of presentation for different subsets of

listeners. In one test the stimuli consisted of the isolated noise intervals, in a second the post-noise intervals alone, while in the third the stimuli were the syllables as recorded.

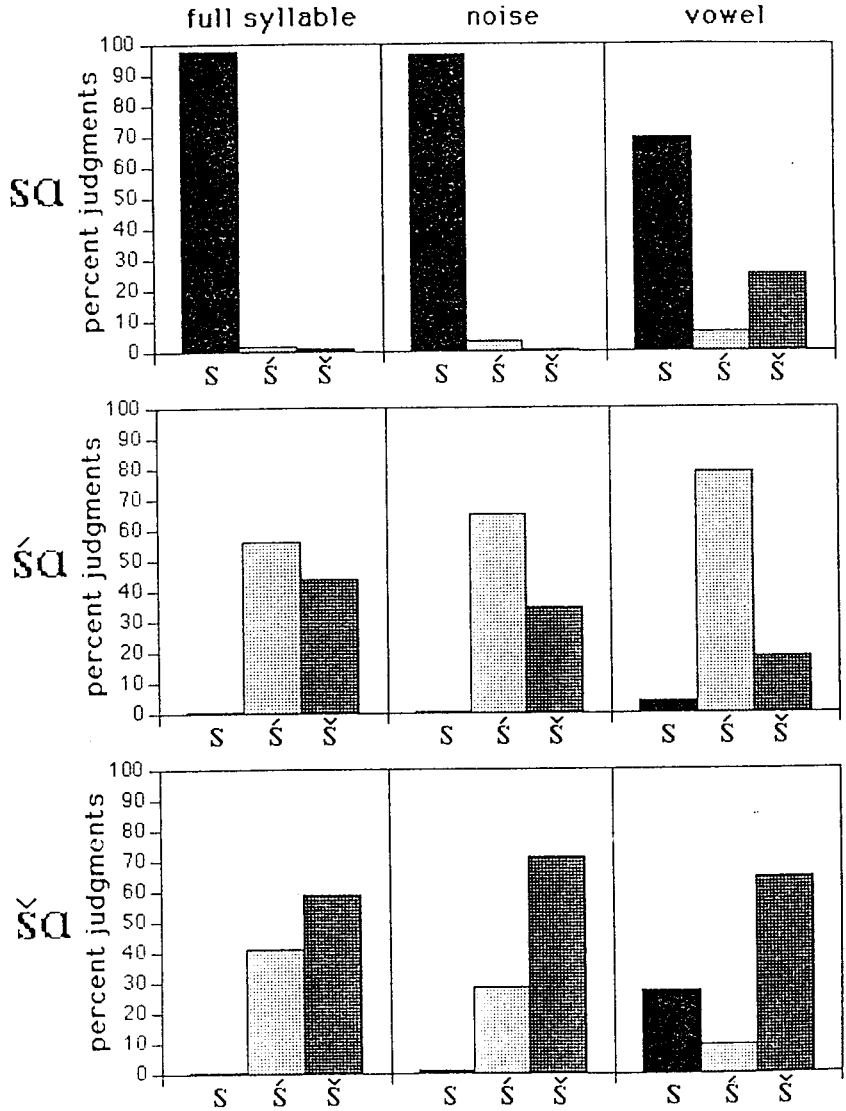

Figure 3. Mean labeling responses of 10 listeners to [sa], [ša] and [šǎ] syllables and their noise and vocalic residues presented separately. A total of 320 judgments (10 Ss × 32 trials) was recorded for each stimulus type.

3.2 Test results

Figure 3 represents the averaged responses for the three tests.[5] Identifications of the three CV stimuli were much as expected: [sɑ] was clearly distinct from both [śɑ] and [šɑ], but the latter two were not nearly as well separated. Thus the slight preference for ś as against š responses in the labeling of the [śɑ] stimuli (56% vs 44%) was not statistically significant (by one sample t-test: $t = .802$; $p = .44$). The slightly greater tendency to report [šɑ] as š rather than ś, 59% vs 41%, was also non-significant ($t = 1.279$; $p = .23$). Moreover, comparison of the responses to the two syllables by paired t-test revealed no significant difference in response patterns ($F(1,9) = 2.1$, $p = .16$). A similar comparison of the responses to the isolated shibilant noises yielded rather different values ($F(1,9) = 12.7$, $p = .0002$), indicating a considerably enhanced ability of our listeners to separate the two. Of the third set of stimuli, in which the vocalic intervals were presented alone, a comparison of responses to those originally preceded by the alveolo-palatal [ś] vs those originally preceded by [š], showed the greatest degree of resolution ($t = 44.1$, $p < .0001$).

4. Auditory difference judgments

In a recently advanced version of the motor theory of speech perception by Liberman and Mattingly (1985), it was suggested that the listener interprets the acoustic signal directly as "intended" articulatory gestures, with no intermediate stage in which its auditory qualities are subjected to cognitive evaluation. The classical case again is that of the stop consonant as cued by formant transitions, whose movements up or down in frequency cannot normally be perceived in any auditorily plausible way,– a transition is not perceived as a shift in pitch or pitches, but as a vocal tract movement to or from closure at a particular place. Since during the interval in which the vocal tract assumes a shape characteristic of a fricative an intelligible acoustic signal **is** produced (so that in this respect it is more like a vowel than a stop), I thought it worthwhile to learn, by means of further experiments, whether the Polish fricative-vowel syllables could be as well or even better separated on the basis of purely auditory judgments than in terms of pho-

netic labelings. In the experiments described below I confined the stimulus set to the two Polish shibilant-vowel syllables.

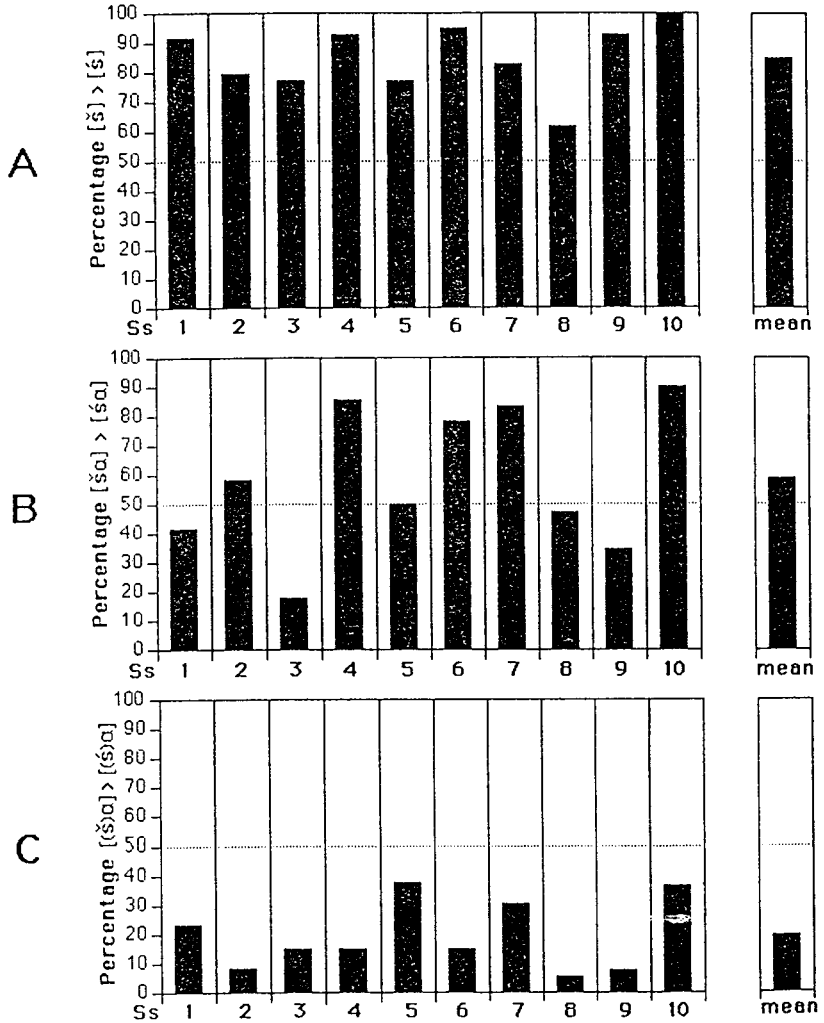

Figure 4. Comparative loudness judgments provided by 10 Ss, each of whom gave a total of 96 responses for each comparison. (A) mean percentage of judgments that isolated [š] noise was louder than [ś] noice. (B) mean percentage of judgments that noise onsets of [ša] syllables were louder than those of [śa], and (C) mean percentage of judgments that residual vowels of [ša] had louder onsets than those of [ś].

4.1 Relative loudness judgments

In one set of tests the listeners' task was to judge the relative loudness of the palato-alveolar as against the alveolo-palatal noises. The task involves the paired comparison of each of the four tokens of the palato-alveolar with each of the four tokens of the alveolo-palatal noises. In one of these tests the vocalic intervals were deleted from the original fricative-vowel syllables. The responses elicited by the isolated noises are shown in panel A of Figure 4, from which it appears that for every one of the ten listeners the palato-alveolar noise [š] was more often than not reported to be louder, a result that might be anticipated from a comparison of both the waveforms and the amplitude curves shown in Figure 1. The mean percentage of judgments favoring [š] as louder (85 %) departs significantly from random ($t = 9.5, p < .0001$). In a related test listeners were asked to compare these noises in their original contexts. Here the results (panel B) are very different; only five of the ten listeners reported the same noise to be the louder one, the other five either went the other way or showed no preference. (By one-sample t-test, the mean percent judgment reporting [š] as louder (58.5 %) is not significantly different from random: $t = 1.10, p = .30$.) The same comparison test was applied to the post-fricative intervals, listeners being asked to decide which one of a vowel pair had the louder onset. From panel C it appears that every one of the ten listeners heard the vowels extracted from [šɑ] as having weaker onsets (less loud) than those derived from [śɑ], again a result very different from random ($t = 8.14, p < .0001$).

4.2 Comparative pitch judgments

Another set of tests, involving the same paired comparisons, asked for relative pitch judgments. Figure 5 shows that for nine of the ten listeners the alveolo-palatal noises ([ś]), both in and out of their original vocalic contexts, had the higher pitch. Again, there were contextually related differences. In the case of the full syllables (panel A) 68 % of the judgments reported [śɑ] as having higher-pitched onsets, and by single sample t-test this score differed significantly from random ($t = 2.4, p = .037$). For the isolated shibilant noises (panel B) the mean percentage of judgments favoring [ś] as higher in pitch was 84.6 %, for which t =

4.2 and p = .002. Moreover, the post-/š/ signals in isolation (panel C were also judged (79.8 %) to begin with the higher pitch (t = 9.6, p .0001).

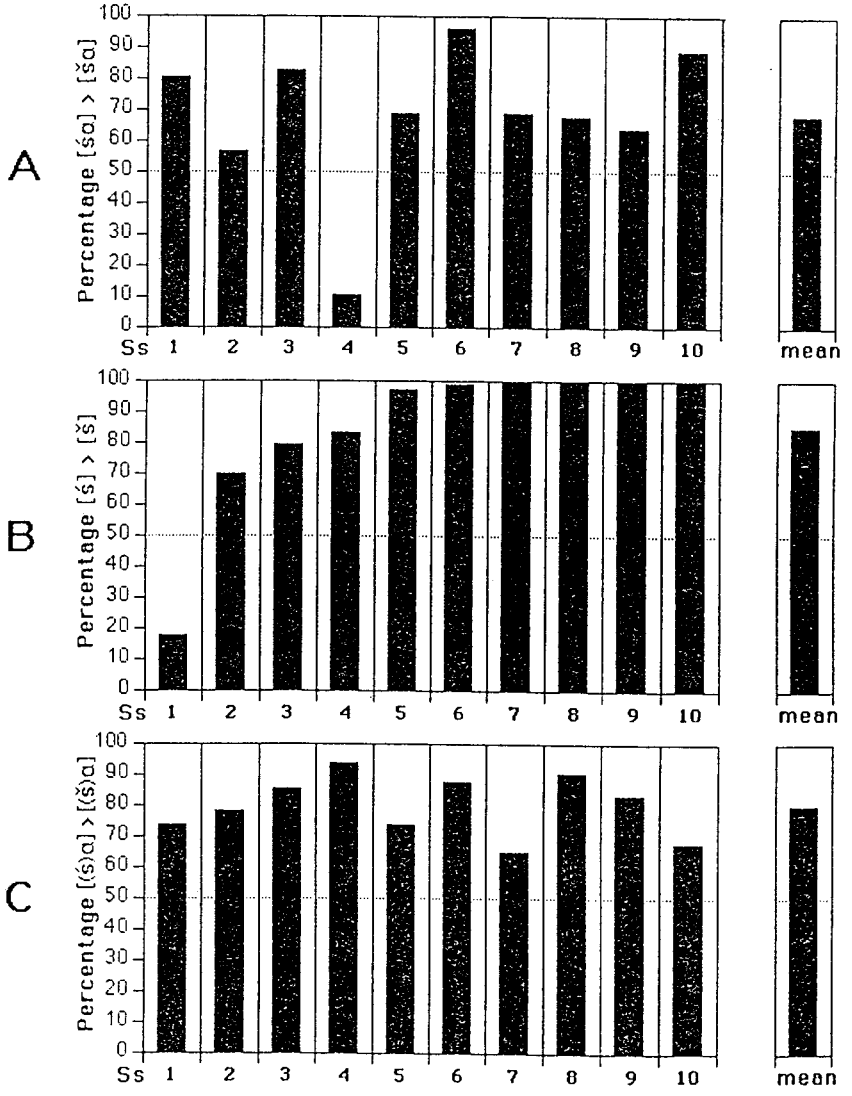

Figure 5. Comparative pitch judgments of 10 Ss, each of whom recorded 96 judgments per comparison. (A) judgments of full syllables, (B) judgments of isolated fricative noises, and (C) judgments of onsets of post-fricative residues.

5. Conclusions

From the identification tests it is clear that the Polish syllable [sɑ], whatever its phonetic differences from English [sɑ] might be, was readily identified as sɑ, but that test subjects did not reliably separate [śɑ] from [šɑ]. Their ability to distinguish [sɑ] from the other two appears to be largely a matter of identifying the [s]-noise, although the post-[s] signal may also have contributed something to its identification. But in the case of [śɑ] vs [šɑ] the failure to apply the ś and š labels consistently is not matched by a similar inability to separate either the noise or the vocalic components of the two syllables, when these were presented separately. The distributions of labeling judgments of both the isolated [ś] and [š] noises show a significant bias in favor of the correct responses, while the post-[š] and post-[ś] intervals were even more successfully identified. It is not immediately obvious why the two shibilant noises should be better identified when the following vowels with their information-bearing transitions have been deleted, or why deleting the initial shibilant noises should yield residues that were even better separated perceptually. To be sure, the observation that in English there is no phonological contrast like that of the Polish shibilant distinction must have relevance to the finding that English-speaking listeners fail to sort the Polish syllables [śɑ] and [šɑ] into distinct categories, as does the fact that both shibilants may be identified more or less equally (?) with English [š]. These two facts, when taken together, might be taken to serve as an adequate explanation of our listeners' failure to distinguish between the [śɑ] and [šɑ] syllables (cf. Best 1995). However, given the absence of an [ś] – [š] distinction in English, how do we then explain their success in labeling the isolated [ś] and [š] noises? Is it plausible to suggest that these signals (though not the full syllables, and surely not the isolated vowels[6]) were perceived psychoacoustically, i.e., that despite the ostensibly phonetic nature of the labels applied to them, the so-called speech mode of perception (Liberman and Mattingly 1985) was not brought into play in the performance of the labeling task required? In order to advance such an argument it would seem wise to find other evidence to support the view that isolated fricative noises are less speechlike than either the vocalic intervals or the full CV syllables. Perhaps the failure to separate the full syllables on the basis of fricative loudness is more prosaically explained by the

fact that while the [š] noise was found to be louder than the [ś] noise, the post-[š] vowel was judged to have a weaker onset than the post-[ś] vowel, and in comparing the full syllables listeners were unable to focus attention strictly on the noise onsets. Moreover, in the case of the pitch judgments, the fact that the syllable [śɑ] was uniformly heard to begin at a higher pitch than [šɑ] is perhaps to be accounted for by the fact that **both** the isolated [ś] noise **and** the post-[ś] vowel onsets had the higher pitch.

Even if it is supposed that in their loudness and pitch judgments listeners were making psychoacoustic rather than phonetic evaluations of [śɑ] and [šɑ] and their components, it is still not clear that this forces us to accept the view that in the process of perceiving speech **as speech** there is a purely auditory stage. For it might reasonably be claimed that fricative noises, particularly in isolation, are no longer unequivocally speech, i.e immediately identified as the outputs of a human vocal tract in particular articulatory configurations, and are therefore amenable to the same psychoauditory processing as any other nonspeech acoustic signals. Of course, we have to be careful in trying to draw a sharp line between phonetic and auditory processing, since after all some phonetic distinctions drawn, e.g the one between "strident" and "nonstrident" fricatives and affricates that many phonologists draw (Chomsky and Halle 1968: 29) would suggest that auditory criteria **can** serve in phonetic classification, though one might wonder whether what the linguist is doing here has any connection with the usual processes of speech perception. The fact that the ś and š noises were more successfully separated on the basis of loudness when presented in isolation than when given in a speechlike context might be construed as evidence for the claim that processing acoustic signals as speech is very different from their auditory evaluation.[7] At the same time, however, we cannot at present deny absolutely that the auditory properties of pitch and loudness play any role in the phonetic classification of these fricative consonants, perhaps particularly by listeners for whom the difference is not a feature of their native language.

Acknowledgment

An earlier version of this paper was presented at the meeting of the Linguistic Society of America held in New Orleans, December 27-30, 1988. I want to thank my colleague Henry Hiż for providing the samples of Polish speech materials used in this study, and for sharing his linguistically informed native-speaker intuitions about the nature of the sounds under examination. Support for this research was provided by NIH Grant HD-01994 to Haskins Laboratories.

References

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. Speech Perception and Linguistic Experience, ed. by V. Strange, 171-204. Baltimore: York Press.

Chomsky, Noam and Morris Halle. 1968. The Sound Pattern of English. New York: Harper and Row.

Delattre, Pierre. 1965. Comparing the phonetic features of English, German, Spanish and French: an interim report. Heidelberg: Julius Gross Verlag.

Harris, Katherine S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech 1.1-7.

Jassem, Wiktor. 1964. A spectrographic study of Polish speech sounds. In Honour of Daniel Jones, ed. by David Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, and J.L.M. Trim, 334-48. London: Longmans.

Ladefoged, Peter. 1967. Stress and respiratory activity. Three Areas of Experimental Phonetics, 1-49. London: Oxford University Press.

Liberman, Alvin. M. and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. Cognition 21.1-36.

Shenker, A. M. 1973. Beginning Polish. New Haven: Yale University Press.

Strevens, Peter. 1960. Spectra of fricative noise in human speech. Language and Speech 3.32-49.

Notes

1 Thus fricatives compared across different languages, though similarly spelled in IPA transcription, may be auditorily distinguished. For example, "[ʃ]" might be said to stand for somewhat different sounds and articulations in English, French, German, Polish and Russian.

2 In IPA transcription these Polish fricatives are represented as [sɕʃ]. The symbols ś ś were used instead of ɕʃ by our subjects because they were found to be more legible as handwritten test responses.

3 Delattre (1965) points out transition differences between English and French [s], which he attributes to differences in tongue position and shaping. Perhaps (?) Polish [s] is more like the French.

4. However, there is no way of knowing from these data whether listeners would fail to distinguish between Polish [s] and English [s], or between Polish [ś] and English [š].

5. The responses of two subjects had to be eliminated from consideration because they reported, and their responses suggested, that they had switched the symbols for the palatoalveolar and the alveolo-palatal fricatives in the course of testing.

6. Under direct questioning, three very experienced linguist-phoneticians who unerringly identified the post-ś signals as against the other two, agreed that the first were **phonetically** distinguishable by the presence of an initial "y" glide.

7 Ladefoged made a similar point in questioning whether estimates of the loudness of isolated vowels were to be taken as judgements of 'heard speech' or of 'meaningless noises' (Ladefoged 1967, p. 39).