Journal of Phonetics (1998) **26,** 345–370
Article ID: jp980079

Crosslinguistic acoustic categorization of sibilants independent of phonological status

Vincent Evers, Henning Reetz and Aditi Lahiri*

Fachgruppe Sprachwissenschaft, University of Konstanz, Fach D 186, 78457 Konstanz, Germany

Received 24 December 1997, revised 18 August 1998, accepted 18 September 1998

The main object of this paper is to provide an acoustic characterization of a stable phonetic contrast across a number of variable dimensions such as vowel context, gender, language, and, in particular, phonological status. The contrast that is investigated is between the dental/alveolar sibilant fricative [s] and its palatoalveolar counterpart [ʃ]. The phonological feature involved in this distinction is [anterior]. Data from three languages are analysed, where the feature has a different phonological status. In English, both fricatives are independent phonemes, and the feature [anterior] is thus contrastive. In Bengali, [s] is an allophone of the phoneme /ʃ/, whereas in Dutch, [ʃ] is an allophone of the phoneme /s/. Power spectra, obtained by placing a 40 ms window in the middle of the friction, display a consistent pattern of differences between [s] and [ʃ] across the three languages independent of gender and vowel contexts. This difference is then quantified by a metric, based on the slopes of the spectral envelope below and above 2.5 kHz. It turns out that all three languages distinguish between [s] and [ʃ] in much the same way, but that the boundary values of the metric show some variation. However, this variation cannot be related to any of the variable factors mentioned above, but seems to be speaker-dependent. It is concluded that phonological status does not affect the realisation of this phonetic distinction, and that the appropriate acoustic correlate displays a relative rather than an absolute kind of invariance.

1. Introduction

Speech researchers tend to assume that the primary unit involved in speech production and perception is the phonetic segment or the distinctive feature, where a word is made up of segments, and the segments themselves are decomposed into a restricted set of phonetic features. Any given feature organises the segments of a language into natural classes, which in turn play a significant role in the phonology. Along with articulatory attributes, a feature is ultimately assumed to be associated with some specific property of the acoustic signal. The precise nature of the relation between such features and these alleged "acoustic correlates" is as yet not straightforward. Many issues are unclear; for instance, should acoustic correlates be thought of as invariant, or as variable according

*Corresponding author. E-mail: aditi.lahiri@uni-konstanz.de

0095–4470/98/040345 + 26 $30.00/0

to phonetic context? Should they generalise across all segment classes? To what extent can the correlates be thought of as language specific? And finally, do the acoustic attributes differ according to their phonological distinctiveness in different languages?

With reference to the last issue, most crosslinguistic studies dealing with acoustic properties have focused on contrasts which are phonemic. Studies which have been concerned with variations within a phonemic inventory have investigated the degree of coarticulation. A possible hypothesis is that speakers try to keep sounds acoustically more distinct when they contrast phonemically rather than when one sound is a phoneme and the other is derived by context, to minimise articulatory effort (Lindblom, 1983). The question of how far phonological constraints influence the degree of coarticulations has been examined, for example, by Manuel (1990). She investigated the F1/F2 vowel space of three Southern Bantu languages: Ndebele, Shona, and Sotho. The languages differ in their vowel inventories; Ndebele and Shona have five vowels [i e a o u], whereas Sotho has seven vowels [i e ɛ ɑ ɔ o u]. Overall her results indicate that Ndebele and Shona speakers raise the vowel /ɑ/ when a higher vowel follows, but Sotho speakers show less coarticulation for /ɑ/ in the same context. She concludes that the lack of coarticulation in Sotho speakers is due to the fact that /ɑ/ has closer neighbours in its vowel inventory than the other two languages. This supports the view that coarticulations are constrained by the phonemic inventory of a language.

To our knowledge, however, there has not been any systematic investigation of an acoustic dimension which is stable across phonetic contexts, but where the feature can lead to a phonemic contrast in one language and an allophonic one in another. For instance, pioneering work by Lisker and Abramson (1964) on the voicing distinctions across 11 different languages focused only on categories which contrast phonemically within each language. Likewise, Lahiri, Gewirth and Blumstein (1984) investigated the acoustic properties of the place of articulation of different diffuse consonants in Malayalam, American English, and French; again, all the contrasts in question involve the same phonemic distinctions. Although some studies do take into account the phonological status of segments, they have compared languages where the segments occur as separate phonemes with languages that have only one of the segments in their inventory. For instance, Jongman, Blumstein and Lahiri (1985) compared the phonemically distinct dental and alveolar stops of Malayalam with alveolar stops of American English and dental stops of Dutch, two languages where the contrast between dental and alveolar stops does not exist. More recently, Utman and Blumstein (1994) have compared the labiodental fricative [f] of Ewe with that of American English. In Ewe, a phonemic contrast exists between labiodental [f] and bilabial [ɸ], while American English has only the labiodental fricative in the labial class. Thus, cross-phonological comparisons found in the literature have not compared languages where a given contrast is distinctive with languages where the same contrast is allophonic, i.e., languages where the same phonetic segments occur with different phonological status.

In this paper, we focus on two segments, [s] and [ʃ], that have precisely this distribution — they are separate phonemes in one language and allophones of the same phoneme in two other languages, /s/ being the phoneme in one language and /ʃ/ in the other. We opted for the sibilant fricatives [s] and [ʃ] because both phonologically and phonetically they form a relatively well-defined and stable class, as indicated by their frequency of occurrence in the world's languages. Maddieson (1984) states that out of a total of 317 languages he investigated, 275 (87%) have at least one of the coronal sibilant phonemes /*s/ or /ʃ/ (where "*s" include both dental and alveolar fricatives).

Within this group, 128 out of 275 (47%) have both /ʃ/ and /s/. Such a high percentage is indicative of the stability of the opposition between these two phonemes.

The sibilants [s] and [ʃ] belong to the class of coronal consonants, where the term "coronal" refers to any sound that is made with the tongue blade (Paradis & Prunet, 1991: 1).[1] As a phonetic categorization, the term "sibilant" is most often isomorphic with [strident] (Keating, 1991: 45), and mainly refers to a relatively large amount of high-frequency noise.[2] In production, strident sounds are characterised by turbulence at the point of articulation due to a supplementary barrier at the constriction (Stevens, 1971). Phonologically, these sibilants are contrasted by the feature [anterior], [s] being [+ anterior] and [ʃ] being [− anterior] (cf. Halle & Clements, 1983; McCarthy, 1988; Lahiri & Evers, 1991, among others). The articulatory correlate of this distinction is the area against the upper teeth or the hard palate where the constriction is made. The acoustic properties of the strident coronals are discussed in detail in the next section.

The three languages we examined are American English, Dutch, and Bengali. In American English, the two sibilants are phonemic (Gussenhoven & Broeders, 1976); in Dutch, the default consonant is /s/ and [ʃ] is derived from it, while in Bengali the opposite relationship holds, such that /ʃ/ is the default consonant and [s] is derived. The goal of this study is twofold: first, to provide an acoustic characterization of this stable phonetic contrast across a number of variable dimensions such as vowel context, gender, and speakers in each of the three languages, and second to investigate whether the same acoustic characterization holds for the three languages in spite of the different phonological oppositions.

1.1. *Previous research*

The acoustic properties of coronal fricatives have been investigated in a large number of studies, including Hughes and Halle (1956), Heinz and Stevens (1961), Strevens (1960), Bladon and Seitz (1986), Clark and Bladon (1986), Behrens and Blumstein (1988), Forrest, Weismer, Milenkovic and Dougall (1988), Nittrouer, Studdert-Kennedy and McGowan (1989) and Nguyen, Hoole, and Marchal (1994). Here we will briefly discuss the main points covered in these papers.

Hughes and Halle (1956) developed a procedure to distinguish between the American English fricatives [s], [ʃ], and [f] on the basis of their gross spectral shapes. The spectra were based on a 50 ms portion of the friction located somewhere in the middle of the fricatives. Since the procedure was intended to apply to voiced fricatives as well, the signal was first high-pass-filtered at 700 Hz, so that any spectral peaks due to voicing would be removed from the spectrum. The metric used to distinguish [s] from [ʃ] was calculated by subtracting the energy (in dB) between 4.2 kHz and 10 kHz from the energy

[1]There is some controversy as to what part of the tongue counts as the blade, and consequently, which consonants have to be counted as coronals (Keating, 1991: 30–31). In particular, it remains unclear whether the palatals should be included as well. Since in this study we are only concerned with the dental/alveolar and palatoalveolar place of articulation, the question of the palatals need not concern us here.

[2]As a phonological feature, [strident] has been used in different forms, depending on the phonological framework. In the framework of Jakobson, Fant and Halle (1952), it was used to formalise distinctions within the classes of labials, coronals, velars, and uvulars. Furthermore, affricates were considered to be strident versions of the corresponding stops. In Chomsky and Halle (1968), the latter usage of the feature was abandoned. In more recent feature-geometric proposals (cf. McCarthy, 1988), [strident] is represented as being dependent on the articulator node Coronal. As such, [strident] can no longer be used to characterize non-coronals.

between 720 Hz and 10 kHz. A small value was characteristic of [s], a large difference indicated a [ʃ]. The metric is essentially a straightforward way to quantify the observation that the frequency peaks for [s] are most often situated above 4 kHz, while the spectra of [ʃ] generally show peaks at lower frequencies. With this metric the authors successfully classified 107 out of 125 (86%) fricative tokens produced by three speakers, both word-initial and word-final, and in the context of front, central, and back vowels. On the basis of these findings, Heinz and Stevens (1961) performed a perception experiment, using synthesised versions of [f], [θ], [s], and [ʃ]. They found that in comparison with the synthesised version of [s], synthesis of [ʃ] required a noise source with a relatively sharp cut-off for frequencies below 3 kHz (Heinz & Stevens, 1961: 592). In the perception experiment, synthesised fricatives were presented to the listeners both in isolation and in prevocalic position. The results indicate that differentiation of [s] and [ʃ] is accomplished primarily on the basis of cues contained in the frication noise itself, whereas the differentiation of [θ] and [f] is accomplished primarily on the basis of the vowel transitions.

Strevens (1960) provides an acoustic analysis of nine English fricatives, including [s] and [ʃ]. The comparison of the spectral characteristics was based on visual inspection of multiple broad-band spectrograms, taken from lengthened fricatives produced in isolation by 13 trained phoneticians. The difference between the alveolar [s] and palatoalveolar [ʃ] turned out to be relatively clear-cut, residing mainly in the lower and upper limits of their noise frequency. For [ʃ], the lowest frequency at which energy was observable varied between 1.6 and 2.5 kHz, but it always exceeded 3.5 kHz for [s]. As for the upper limits, [ʃ] did not show energy above 7 kHz, while the spectrograms for [s] indicated that energy was present above 8 kHz.

In a study of the acoustic characteristics of American English voiceless fricatives, Behrens and Blumstein (1988) investigated spectral differences along with other acoustic parameters. They found that the individual sibilants did not differ systematically in duration and amplitude. Spectral information was measured at three different points in the fricative: at fricative onset, in the middle of the noise, and immediately preceding voicing onset. However, the spectral characteristics turned out to remain relatively stable over the duration of the frication noise. The two sibilants could be reliably distinguished by major frequency peaks between 3.5–5 kHz for [s] and 2.5–3.5 for [ʃ]. Behrens and Blumstein conclude: "… any generalized patterns based on spectral properties can probably be derived from either a static configuration of the frication noise itself, or of the frication noise relative to the following vowel, irrespective of where the friction noise is measured" (pp. 297–298).

Bladon and Seitz (1986) found that the orientation of the low frequency spectral edge of the first peak of fricative noise was able to discriminate between [s] and [ʃ] independently of vowel context and gender. Their result held only, however, if Bark-transformed spectra were used. In the same vein, Clark and Bladon (1986) investigated the fricatives of the language Shona, where the class of voiceless coronal fricatives has three members: dental/alveolar [s], palatoalveolar [ʃ], and labialised alveolar [s^w]. A perceptual experiment with synthesised fricatives revealed that identification of [s] was dependent on the spectral slope in the higher frequencies in combination with the location of the peak. On the other hand, [s^w] and [ʃ] were differentiated mainly on the basis of the vowel transitions. The spectra of both [ʃ] and [s^w] were characterised by a relatively steep slope in the low frequencies. Again, all spectra were Bark-transformed.

Forrest *et al.* (1988) aimed at establishing a quantitative metric which would provide a classification of word-initial obstruents independent of speaker and vowel context. Their metric is based on a statistical analysis of power spectra which takes into account mean, skewness, and kurtosis. Classification of the different obstruent categories was achieved by means of a discriminant analysis of these three spectral moments. It turned out that discrimination was particularly successful for the coronal sibilants (based on the moments for the first 20 ms of the consonants), which were correctly classified as [s] or [ʃ] in about 98% of the cases. However, since the recordings were taken from a corpus originally collected for different purposes, the data set was limited to only two words, *see* vs. *she*, which were read six times by five male and five female speakers. In accordance with the findings of Bladon and Seitz (1986), it was found that the skewness of the Bark-transformed spectra was the most important property distinguishing [s] from [ʃ] (ibid: 122). On closer inspection it turns out, however, that the Bark scale only improved classification for the male speakers (80.0% versus 98.3%). For the female speakers classification rates actually decreased from 100% to 94% when the Bark transformed spectra were used. These results thus fail to provide conclusive evidence regarding the use of Bark spectra in the analysis of fricatives. In fact, this issue still awaits a systematic investigation.

Nittrouer *et al.* (1989) investigated the contrast between voiceless alveolar and palato-alveolar fricatives and the amount of fricative-vowel coarticulation in productions of children and adults. They found that the so-called "centroid", or center of gravity of the DFT spectra (the first moment of the spectral distribution), reliably distinguished between the spectral shapes of the two fricatives (ibid: 124), yielding higher values for [s] than for [ʃ]. This analysis is comparable to that of Forrest *et al.* (1988), but the latter obtain greater detail and classify more obstruent distinctions by taking into account higher spectral moments as well.

Nguyen *et al.* (1994) aimed at establishing a statistical model which would permit the regeneration of spectra of [s] and [ʃ] on the basis of a limited set of articulatory parameters. To that end they performed a principal components analysis on simplified Bark-transformed spectra of French dental and palatoalveolar fricatives. The first three factors of the principal components analysis turned out to be important in distinguishing the sibilants. The relationship between the factors and the articulatory data was then investigated by means of multiple linear regression. Finally, the output of this procedure was used to regenerate fricative spectra. A comparision between the regenerated and the original spectra revealed a high level of accuracy. The authors conclude that the tongue has relatively few degrees of freedom, since spectra of [s] and [ʃ] can be regenerated on the basis of only a few articulatory parameters. As for the acoustic characterization of the sibilants, although the outcome of the principal components analysis could be correlated with specific patterns of energy distribution in the spectra, the authors do not discuss this relationship any further.

The papers mentioned above allow us to draw a number of conclusions. First, it turns out that amplitude and durational properties do not contribute to the differentiation of the coronal sibilants (Behrens & Blumstein, 1988). Second, several studies have suggested that the sibilants [s] and [ʃ] are differentiated by the spectral properties of the friction itself, rather than by any information from the vowel context (Hughes & Halle, 1956; Bladon & Seitz, 1986; Clark & Bladon, 1986). Third, it has been consistently observed that there is more low-frequency energy for the [− anterior] [ʃ] than for the [+ anterior] [s].

Certain questions, however, remain unanswered. Since almost all the studies focused on a single language — English in most cases, where the contrast between the two sounds is phonologically distinctive—the possible influence of phonological status is not addressed. Second, with the exception of Hughes and Halle (1956), the sounds have been contrasted only in word-initial position. Third, the appropriateness of using the Bark scale for analysis of fricatives is not entirely obvious. Since the Bark scale separates prominent spectral peaks, it is successful for vowel analysis, but it is not necessarily as suitable for discrimination of fricatives. For example, as we mentioned before, although Forrest *et al.* (1988) found that the use of the Bark scale improved the overall classification of the fricatives, the discrimination of [s] from [ʃ] improved only for the male speakers. They also found a better discriminability for some sounds with the linear scale and for other sounds with the Bark scale (ibid:122).

Finally, although it is obvious from the data that an appropriate acoustic analysis should take into account spectral shape as a reflection of the characteristic energy difference between the low and the high frequencies, several ways have been proposed to implement this observation. The traditional studies have considered differences between absolute energy values in two frequency regions (cf. Hughes & Halle, 1956). The more recent studies apply general statistical procedures to the spectra. The present study adopts an analysis which is intermediate between these two positions. We aim at integrating all the relevant information from earlier research in order to investigate the acoustic properties of [s] and [ʃ] in three languages where they differ in phonological status. On the basis of the finding that the distinction between [s] and [ʃ] is situated in the frication noise rather than the vocalic context, we have decided to concentrate on the noise portion of the signal. In addition, instead of comparing absolute energy values in two frequency regions, or using gross statistical methods to separate data points, we focus on spectral slopes within the low and high frequency range and provide a metric that is successful in classifying these sibilants across languages and phonologies.

The analysis we propose is a straightforward quantification of the spectral slopes within the low and high frequency range, and provides a metric that is successful in classifying the sibilants across languages and phonological inventories. The reason for using spectral slopes is grounded on both articulatory and acoustic considerations. For sibilants, the back cavity hardly contributes to the spectral slope (Stevens, 1989:23). Rather, the resonances of the vocal tract in front of the constriction determine the frequencies of the spectral peaks. With a more anterior constriction, as for [s], the front cavity is smaller and hence the resonance frequencies are higher. In contrast, for [ʃ], which has a more posterior constriction, the resonance frequencies are somewhat lower. Since the sibilants have a noise source, their spectrum does not necessarily show a clear formant structure (Fant, 1960). Hence, spectral slope is a good candidate to capture the differences in the location of the resonance frequencies which we expect will differentiate the sounds based on their constrictions.

2. Method

2.1. *Languages*

This research investigates the phonetic contrast between [+ anterior] and [− anterior] sibilants in languages where this contrast has a different underlying phonological status.

We have chosen English, Dutch and Bengali, as each of them represents one of the three relevant phonological contrasts. In English, both /s/ and /ʃ/ are separate phonemes. In Bengali, [s] is an allophone of /ʃ/, whereas Dutch has the reverse situation, [ʃ] being an allophone of /s/.

Some details concerning the distribution of these fricatives in the three languages are necessary to justify the choice of words. In English, both consonants occur in word initial, medial, and final positions: *see*, *she*; *missive*, *mission*; *gas*, *gash*. The only distributional constraint in English is that in consonant clusters [s] is widely permitted and not [ʃ]: *spin*, *stack*, *bask*, etc. In Bengali, the default sibilant is [ʃ], and the dental fricative [s] occurs in word initial consonant clusters and word medially only before dental stops:[3] [spɔrd̪a] 'audacity', [skɔnd̪^{h}o] 'shoulder', [mɔst̪o] 'huge', [kust̪i] 'wrestling', etc. (Lahiri, 1991). The clearest examples of the assimilation of the sibilant to the following dental stop are found in verbal paradigms with different affixes: cf. the verb 'to sit' [boʃ-i] 1st person present, [boʃ-lo] 3rd person future, [boʃ-bo] 1st person future, but [bos-t̪o] third person past. In all other positions, Bengali has [ʃ]: [ʃolo] 'sixteen', [mɔʃɑ] 'mosquito', [maʃ] 'month', [mɔʃla] 'spices'. In contrast to Bengali, the default consonant in Dutch is [s]. The palatoalveolar sibilant [ʃ] occurs sporadically in loan words such as *shoarma* 'Turkish meat dish' and *quiche*, but it is most frequently found as an allophone of /s/ in front of the palatal glide [j]. This assimilation occurs across words as well as word-internally. It is most common in the diminutive forms of words ending in [s]: *ijs* [ɛɪs] 'ice cream' but *ijsje* [ɛɪʃə] 'a small ice cream'. The /s/ assimilates to the [− anterior] property of the diminutive affix which begins with [j], a palatal glide.[4]

2.2. *Speech material*

Since in Dutch and Bengali, [s] and [ʃ] alternate only in word medial position, for each language a set of near minimal word pairs was chosen with the contrasted sibilants in postvocalic position. The choice of the preceding vowels was necessarily restricted by the inventory of each language. For every language we selected prototypical rounded and unrounded front and back vowels varying in height. In Dutch and English, there exists an additional distinction between tense and lax vowels, and we opted for the lax vowels since intervocalic [s] after tense vowels is extremely rare in Dutch. In all we had four lax vowels for Dutch, [ɪ œ ɔ ɑ]; four vowels in Bengali, [i u ɔ ɑ]; and six lax vowels for

[3]Word initial clusters are rare in Bengali and [ʃt̪] clusters are generally found in learned words.

[4]There are various reports regarding the precise place of articulation of the [− anterior] fricative in the three languages. Keating (1991: 41) states that, in general, the boundary between the two values of [anterior] is situated at the midpoint of the alveolar ridge, between the upper teeth and the corner of the ridge. For Dutch, Booij (1995: 7, 95) describes the [− anterior] fricative as a postalveolar sound derived from a sequence of [s] and [j] and transcribes it as [ʃj] as in [pɑʃja] 'pasha'. And Collins and Mees (1984: 160) say the following: "The sequences /sj, zj/ are realised as alveolo-palatal fricatives [ɕ ʑ]. They are articulated with an extensive area of the blade and back of the tongue, raised to form a light contract with the rear of the alveolar ridge and the forward portion of the hard palate … . D(utch) /sj/ has sharper friction than the English /ʃ/ and generally is more obviously palatal …." Further, English is claimed to have lip rounding in the production of the palatoalveolar fricatives as opposed to Dutch (Ladefoged, 1982: 59; Gussenhoven & Broeders, 1976: 115). As for Bengali, Chatterji (1975: 546) states that it has a palatoalveolar [ʃ] and describes it as being a more palatal sibilant than the dentalised articulations in the neighbouring languages. What is important for us is that there is always a two-way contrast in each of the three languages and that there is only one [− anterior] fricative which is articulated behind the alveolar region. Our goal is to see if the contrast is acoustically manifested in the same way in the three languages. For the sake of simplicity we will refer to the [− anterior] sound as a palatoalveolar [ʃ].

English, [ɪ ʊ ɛ ʌ ɒ æ]. Dutch does not have a high back rounded vowel which is lax; therefore, we had to take the closest front counterpart [œ]. One reason for using more vowels in English was our endeavour to match them with both Dutch and Bengali. Since many American English dialects round the back low vowel, we included both [æ] and [ɒ]. The vowel [ɛ] was included to have an extra non-low front vowel to match Dutch [œ]. Finally, since we could find only one reasonable pair of words with [ʊ] in English, we also included the phonetically closest lax vowel [ʌ].

The environment following the fricative was kept as similar as possible. In Dutch, the sibilant was always followed by a [ə], and in English the following vowel was either a [ə] or an [i]; in Bengali, since [s] occurs medially only before dental stops, the words with [ʃ] had a consonant following the fricative as well. In total, there were 12 pairs of words for each language. The complete list of words, which includes the phonetic values of the vowels preceding the fricatives, is given in the appendix.

2.3. *Procedure*

For each language, each word pair was read three times by two female and two male speakers from randomised lists. The recordings were made on a DAT-recorder (Sony TCD 1000) using a Sennheiser MD 211 N microphone, set at flat spectral position. A total of 864 words were then digitised into the computer with a sampling rate of 20,000 Hz, and subsequent analyses were made using a speech waveform analysis program developed by the second author.

A power spectrum was computed for each token using a 40 ms window placed at the middle of the friction. The choice of the window size and the position of the window was based partly on the studies reported above and on perceptual results. Jongman (1989) carried out a perceptual study on English voiced and voiceless fricatives, investigating the duration of frication noise needed for correct identification of place of articulation, manner (fricatives as opposed to stops), and voicing. For coronal sibilants, he found that the first 40 ms of frication noise was sufficient for listeners to identify [s] with 71% accuracy and [ʃ] with 89% accuracy. An additional 30 ms of frication increased the identification scores for [s] and [ʃ] to 82% and 98% respectively. A 100% identification was possible when listeners heard the complete frication. Compromising between the fact that the first 40 ms of frication noise was sufficient for 71% identification of [s] while the complete frication was necessary for 100% identification, we chose a 40 ms window placed in the middle of the frication to ensure that we had sufficient information to disambiguate the two fricatives.

Each 40 ms window consisted of 800 signal samples, which were extended with 224 zero values, giving a total of 1024 data points. A power-spectrum was computed from the 1024 point FFT-spectrum. The output was a 512 point frequency spectrum indicating the energy distribution of the 40 ms speech segment from 0 to 10 kHz.

Fig. 1 shows typical waveform patterns of two sibilants with the location of the window. The window was placed on the basis of auditory and visual inspection of the waveform.

Comparing the spectra for [s] and [ʃ] across the differential languages, we observed that within a single time frame, the slope of the spectral envelope in two frequency regions—below and above 2.5 Hz—characterised the main distinction between the fricatives, in accordance with the articulatory evidence. From the visual inspection it seemed obvious that the sibilants could be distinguished if the observed spectral slopes

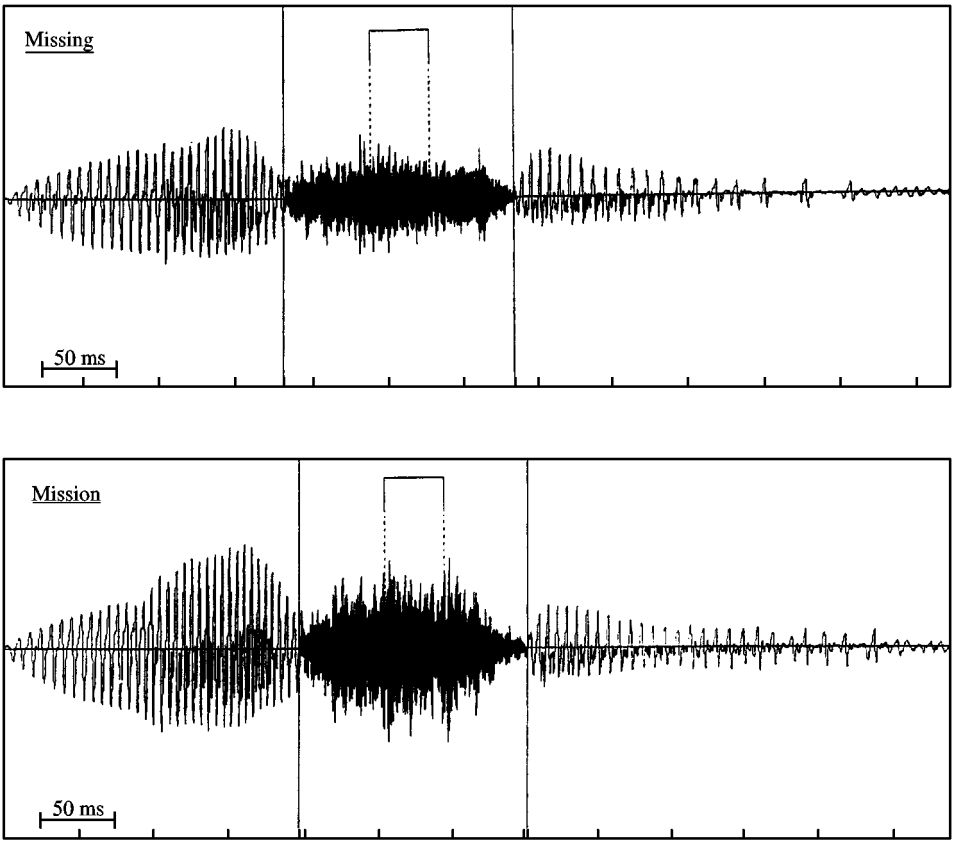


Figure 1. Typical waveform patterns of two words containing the sibilants in intervocalic position: English *missing* and *mission*. The vertical lines indicate the onset and offset of the fricative. The location of the spectrum window for each fricative is indicated on top of the waveform.

could be reduced to two straight lines—one below 2.5 kHz and the other above that frequency region up to 8 kHz. Above 8 kHz, the spectra did not show any distinguishing characteristics.

There is a further reason for considering the global spectrum rather than specific formants. These sibilants have one sound source at the tongue constriction and another source at the teeth, where the sharp airstream from the tongue constriction creates a turbulence (Fant, 1960: 178). It is rather difficult to predicate absolute formant locations and the formant peaks themselves can be attenuated by additional zeros which are part of the noise source spectrum. Overall, the spectral shape of a fricative is likely to be more flat than that of a vowel, which normally shows a clear formant structure. Therefore, it is preferable to describe the spectrum of the sibilants in terms of spectral shapes rather than formant patterns (Fant, 1960: 26).

To examine the distribution of intensity and frequency of the spectra, we computed one linear regression line for the values below 2.5 kHz and a second for all values between 2.5 kHz and 8 kHz. For both regression lines, the intensity values were treated as dependent on the frequency values. Thus, for each spectrum, two coefficients were

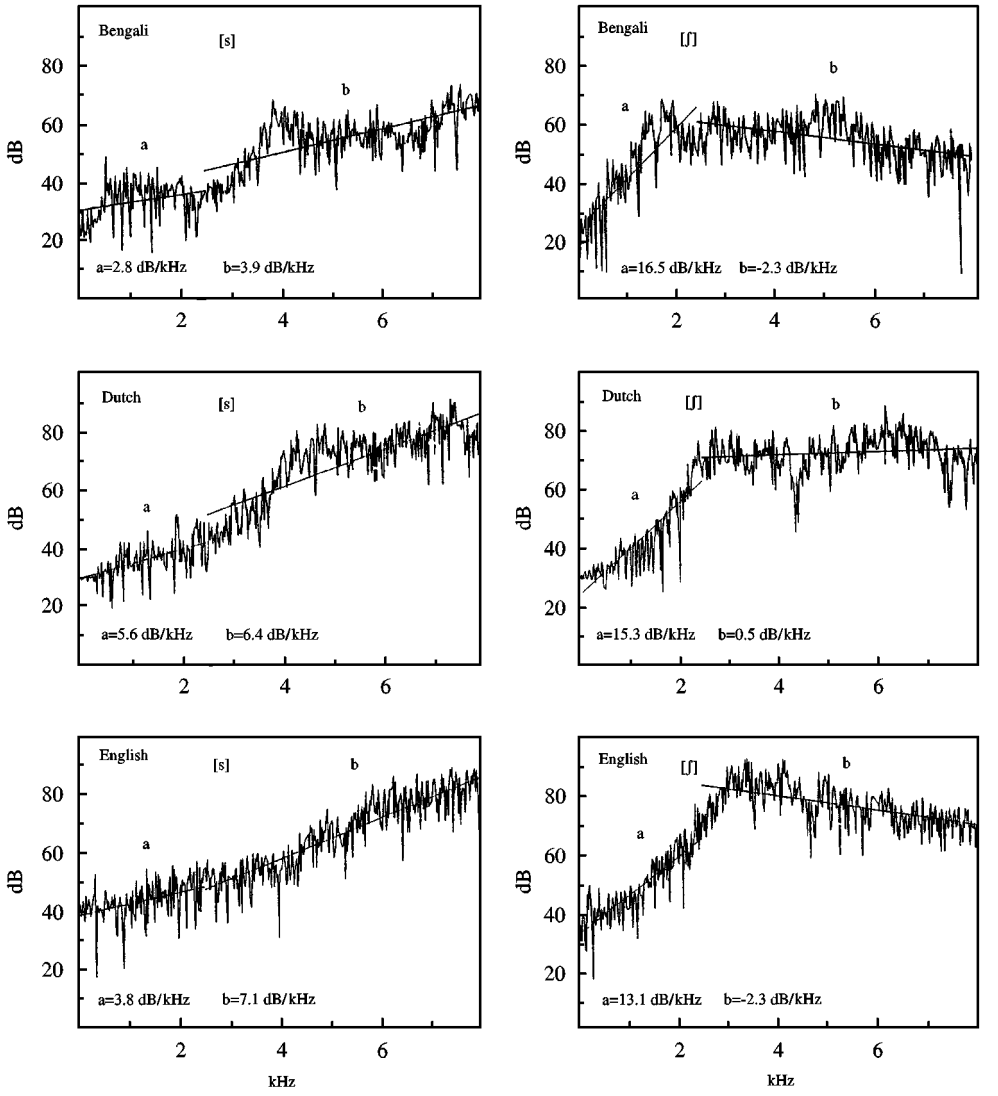


Figure 2. Spectra with superimposed regression lines taken from one word pair in each language. The regression lines of each spectrum are labelled **a** and **b**, and the steepness of both **a** and **b** are included in each box.

obtained: one giving the steepness of the regression line from 0 kHz to 2.5 kHz, and one from 2.5 kHz to 8 kHz, the unit of this spectral steepness being dB/kHz. These regression line coefficients served as input to the subsequent analyses. Fig. 2 gives examples of spectra with the two regression lines. The [s] spectra are in the left hand column and the [ʃ] spectra on the right, one pair for each of the three languages. The superimposed regression lines emphasise the general pattern we described above, viz., that the slope of the spectral envelope differs in the low and high frequency regions for the two sibilants.

3. Results

In the following presentation of the results, we will refer to the individual speaker by means of abbreviations. The language code comes first (B = Bengali, D = Dutch, E = English), gender second (F = female, M = male) and speaker number third (1, 2). Thus, for example, BF2 refers to the second female Bengali speaker.

The regression lines quantify the orientation of the spectral envelope within two prominent frequency regions. To examine the relation between the steepness of the two regression lines, we graphically represented the data in two-dimensional scatterplots (Fig. 3), where **a** denotes the steepness below 2.5 kHz and **b** denotes the steepness above 2.5 kHz. Each data point in a scatterplot represents a single spectrum, with the steepness of the low frequency regression line **a** plotted along the y-axis, and the steepness of the high frequency regression line **b** plotted along the x-axis. Spectra of [s] are plotted as filled circles, and spectra of [ʃ] as open circles. Fig. 3 gives a scatterplot for each speaker. As can be seen, there is a clear and consistent separation of [s] and [ʃ] within and across subjects. The generalisation is obvious—the clustering of the two consonants is always the same, [ʃ] being located in the left top quadrant, while [s] is always below and somewhat to the right. Their relative position indicates that the left (low frequency) regression line for [ʃ] (below 2.5 kHz) is always steeper than that of [s], whereas the right (high frequency) regression line for [ʃ] is either the same or less steep than that of [s]. The location of the boundary distinguishing the sounds in each of the languages is, therefore, anything but arbitrary. There are no speakers for whom the boundary is to be found, for example, in the bottom right corner of the quadrant. This means that, given the nature of our analysis, the relative positions of the spectra of [s] and [ʃ] in the quadrant constitute the property that distinguishes the two sibilants.

Although for each individual speaker the two sibilants occupy distinct regions on the plane, a comparison across speakers reveals that there is a certain amount of speaker-specific variance in the location of the clusters. That is, between individual speakers there are differences with regard to the distance between the clustering of [s] and [ʃ]. This is not language specific but holds for all three languages. For instance, [s] and [ʃ] are equally far apart for the Bengali speaker BM1 as for the Dutch speaker DF1. Although English shows the clearest categorization overall, in this language there are speaker-dependent differences as well. In particular, EM1 shows a relatively small distance between clusters similar to BF2 and DM2. The variance that we observe appears not to be gender dependent either. For Dutch, DF1 shows the largest amount of variability within each of the two sibilant categories, whereas for Bengali BM2 shows the largest differences within categories. This pattern suggests that clustering differences within sibilant categories are neither gender specific nor language specific. At first glance, therefore, it appears that regardless of the different phonological status of the two consonants in the three languages, they have similar acoustic differences. To establish this more precisely we quantified the acoustic distinctions that we observe in the scatterplots.

Recall that the points in the scatterplots represent the steepness of the two regression lines. The difference between the two sibilants can be expressed in terms of the particular quadrant in which the spectra are clustered. This suggests that it should be possible to separate [s] from [ʃ] by using the *difference* in the rate of increase of spectral energy below and above 2.5 kHz. Since **a** denotes the steepness below 2.5 kHz and **b** denotes the steepness above 2.5 kHz, ideally speaking the relationship $\mathbf{(a - b)}_{ʃ} > \mathbf{(a - b)}_{S}$, expressed in dB/kHz, ought to hold across all speakers and languages. That is, for [ʃ], the spectral

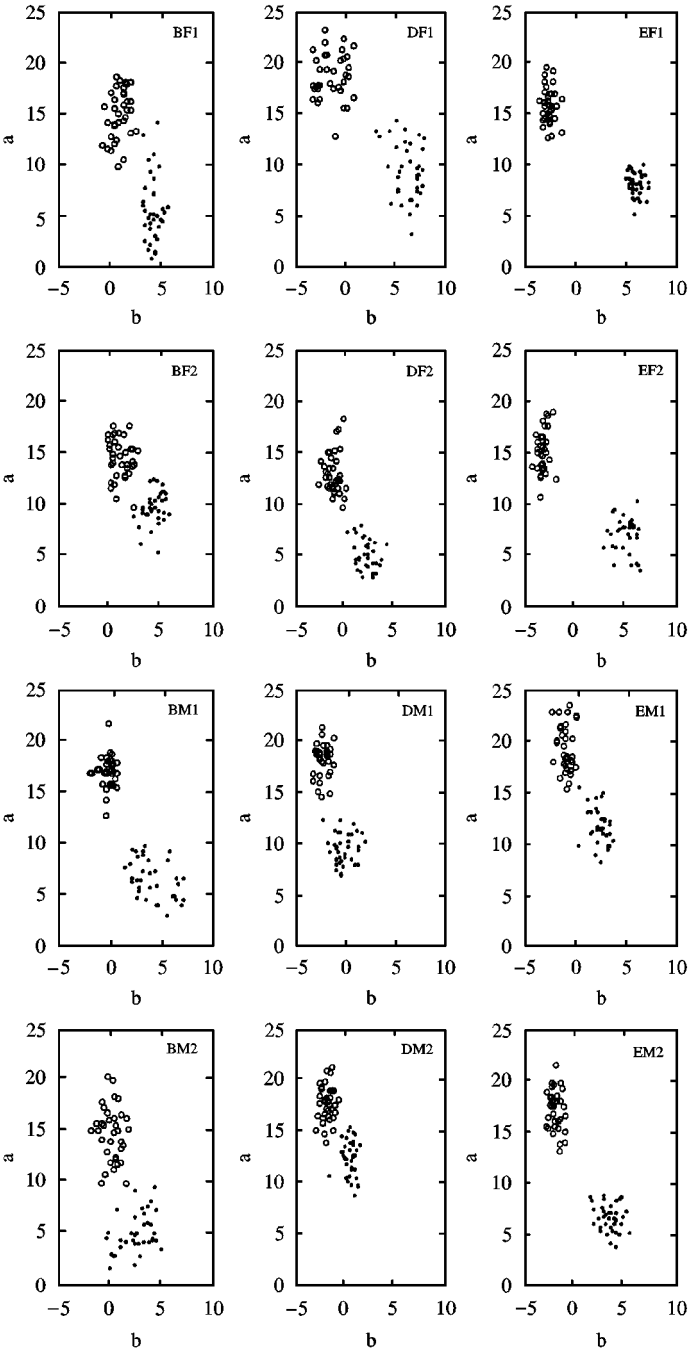


Figure 3. Clustering of the fricative spectra according to the steepness of the two regression lines. Each data point in a scatterplot represents a single spectrum, with the steepness of the low frequency regression line **a** plotted along the *y*-axis, and the steepness of the high frequency regression line **b** plotted along the x-axis. Spectra of [ʃ] are plotted as open circles, spectra of [s] as filled circles. Each scatterplot contains all data points for a single speaker.

TABLE I. Frequency counts of the STEEPNESS DIFFERENCE metric (**a** − **b**) for spectra of [s] and spectra of [ʃ]. The columns represent the respective counts of (**a** − **b**) values, which are indicated on the top line; only the range from 6 to 16 is represented since this is the region of ambiquity. The large top panel contains the frequency values for each individual speaker. The middle panel gives the values for each of the three languages both in absolute numbers and in percentages, and the bottom panel gives the values pooled across all data for the two fricatives

(**a** − **b**)	≤ 6	7	8	9	10	11	12	13	14	15	≥ 16	∑
$BF1_s$	32	2	0	0	2	.	.	.	.	.	.	36
$BF1_ʃ$	.	.	.	2	0	3	3	7	4	4	13	36
$BF2_s$	31	3	2	.	.	.	.	.	.	.	.	36
$BF2_ʃ$	.	1	0	0	1	7	6	4	4	4	9	36
$BM1_s$	34	2	.	.	.	.	.	.	.	.	.	36
$BM1_ʃ$	.	.	.	.	.	.	.	1	0	2	33	36
$BM2_s$	35	1	.	.	.	.	.	.	.	.	.	36
$BM2_ʃ$	.	.	1	0	1	5	4	3	4	3	15	36
$DF1_s$	30	2	0	3	1	.	.	.	.	.	.	36
$DF1_ʃ$	.	.	.	.	.	.	.	.	1	1	34	36
$DF2_s$	34	2	.	.	.	.	.	.	.	.	.	36
$DF2_ʃ$	.	.	.	.	2	1	5	10	3	5	10	36
$DM1_s$	.	3	5	9	5	8	3	2	0	1	.	36
$DM1_ʃ$	.	.	.	.	.	.	.	.	.	.	36	36
$DM2_s$	.	.	3	2	5	4	8	8	3	3	.	36
$DM2_ʃ$	.	.	.	.	.	.	.	.	.	.	36	36
$EF1_s$	36	.	.	.	.	.	.	.	.	.	.	36
$EF1_ʃ$	.	.	.	.	.	.	.	.	.	2	34	36
$EF2_s$	36	.	.	.	.	.	.	.	.	.	.	36
$EF2_ʃ$	.	.	.	.	.	.	.	.	2	.	34	36
$EM1_s$	2	4	3	10	7	2	4	3	0	1	.	36
$EM1_ʃ$	.	.	.	.	.	.	.	.	.	.	36	36
$EM2_s$	35	1	.	.	.	.	.	.	.	.	.	36
$EM2_ʃ$	.	.	.	.	.	.	.	.	.	3	33	36
B_s	132	8	2	0	2	.	.	.	.		.	144
$B_ʃ$	.	1	1	2	2	15	13	15	12	13	70	144
D_s	64	7	8	14	11	12	11	10	3	4	.	144
$D_ʃ$	.	.	.	.	2	1	5	10	4	6	116	144
E_s	109	5	3	10	7	2	4	3	0	1	.	144
$E_ʃ$	.	.	.	.	.	.	.	.	2	5	137	144
B_s%	91.7	5.6	1.4	0.0	1.4	.	.	.	.	.	.	100
$B_ʃ$%	.	0.7	0.7	1.4	1.4	10.4	9.0	10.4	8.3	9.0	48.6	100
D_s%	44.4	4.9	5.6	9.7	7.6	8.3	7.6	6.9	2.1	2.8	.	100
$D_ʃ$%	.	.	.	.	1.4	0.7	3.5	6.9	2.8	4.2	80.6	100
E_s%	75.7	3.5	2.1	6.9	4.9	1.4	2.8	2.1	0.0	0.7	.	100
$E_ʃ$%	.	.	.	.	.	.	.	.	1.4	3.5	95.1	100
$∑_s$	305	20	13	24	20	14	15	13	3	5	.	432
$∑_ʃ$	.	1	1	2	4	16	18	25	18	24	323	432
$∑_s$%	70.6	4.6	3.0	5.6	4.6	3.2	3.5	3.0	0.7	1.2	.	100
$∑_ʃ$%	.	0.2	0.2	0.5	0.9	3.7	4.2	5.8	4.2	5.6	74.9	100

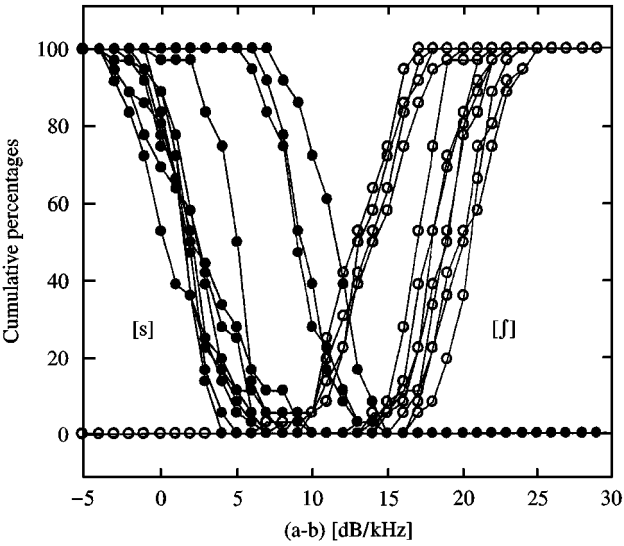


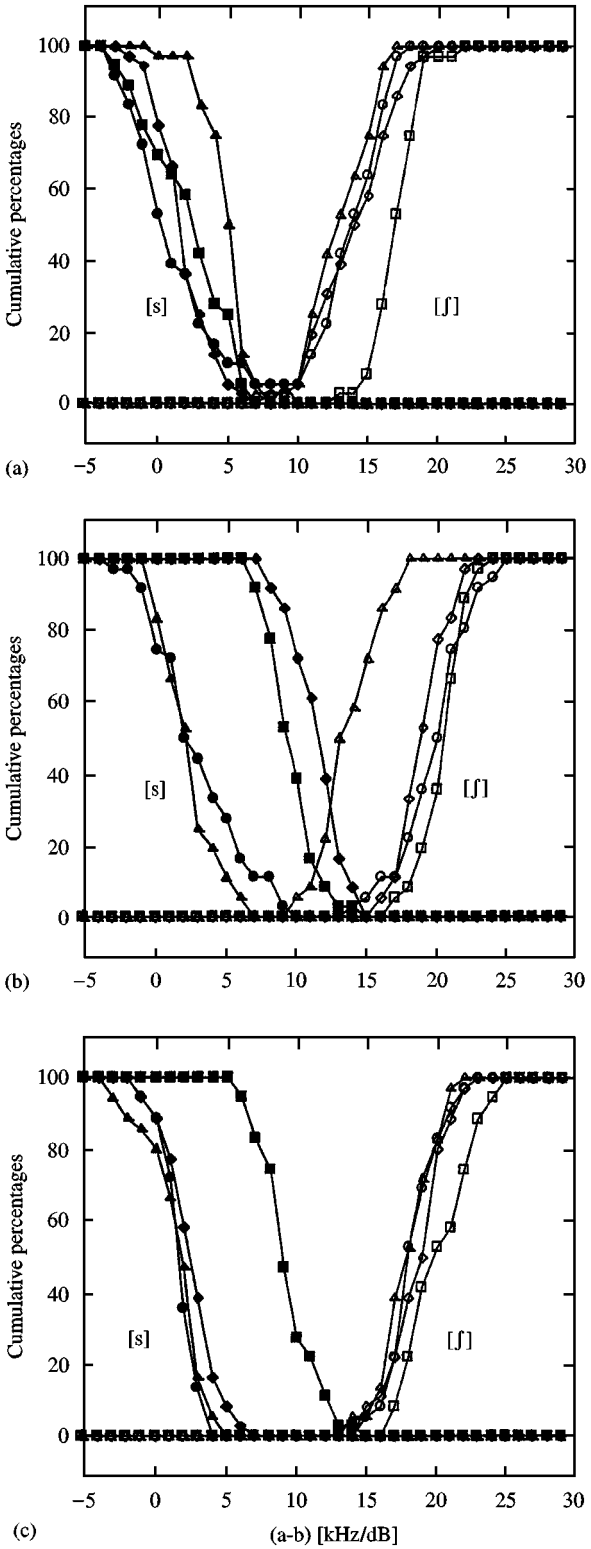
Figure 4. Cumulative percentages of STEEPNESS DIFFERENCE values for all 12 speakers. The x-axis marks the values for $(\mathbf{a} - \mathbf{b})$, i.e. the difference in steepness between the low frequency and high frequency regression lines of a spectrum. Each line represents the cumulative curve of a single speaker for one of the sibilants. Open circles represent spectra of [ʃ], filled circles represent spectra of [s].

energy rises sharply below 2.5 kHz and then the slope drastically levels out to zero or below. For [s], the spectral energy rises much slower in the low frequencies, and there is usually a smaller difference in rate of increase between the lower and the higher frequencies. Note that even if the rate of increase is the same in one frequency range for a pair of stimuli, the steepness in the other range would reflect the difference. That is, if **a** happens to be the same for a pair of sounds, $\mathbf{b}_{\int}$ should be less than $\mathbf{b}_{\mathbf{S}}$, and vice versa, if **b** is the same, the relationship $\mathbf{a}_{\int} > \mathbf{a}_{\mathbf{S}}$ should hold. This measure, which we denote as the STEEPNESS DIFFERENCE measure in db/kHz, is thus able to quantify the observational differences we have noted.

Does the above relationship hold for all speakers and all languages? To test this hypothesis we computed the STEEPNESS DIFFERENCE $(\mathbf{a} - \mathbf{b})$ for all spectra and plotted their distribution in cumulative line graphs. We wanted to determine (a) the range within which speakers differentiated [s] and [ʃ], i.e., within what range the relationship $(\mathbf{a} - \mathbf{b})_{\int} > (\mathbf{a} - \mathbf{b})_{\mathbf{S}}$ would hold for individual speakers, and (b) to what extent this range varies across languages.

The top part of Table I gives the frequency of the STEEPNESS DIFFERENCE values for individual speakers. There are no spectra of [ʃ] for which the value $(\mathbf{a} - \mathbf{b})$ is less than 7,

Figure 5. Cumulative percentages of STEEPNESS DIFFERENCE values for the (a) Bengali (b) Dutch and (c) English speakers. The individual speakers are indicated by means of symbols: (a) circle = BF1, triangle = BF2, square = BM1, diamond = BM2. (b) Circle = DF1, triangle = DF2, square = DM1, diamond = DM2. (c) Circle = EF1, triangle = EF2, square = EM1, diamond = EM2. Open and closed symbols: see legend for Fig. 4.


Cumulative percentages
100
80
60
40
20
0
[s]
[ʃ]
−5
0
5
10
15
20
25
30
(a)
(b)
(c)
(a-b) [kHz/dB]

and there are no spectra of [s] for which (**a** − **b**) exceeds 15. Therefore, all speakers situate the boundary between the two consonants within that range. In order to see the distribution of spectra around the boundary values, we plotted the data in a cumulative line graph (Fig. 4). Each line represents a single speaker, where the filled circles on the line represent [s] and the open circles represent [ʃ]. Although this graph clearly shows that the two sibilants are very well separated by the metric, it can also be observed that there is a certain amount of overlap between the two sibilant categories across speakers. The numbers in the top part of Table I indicate, however, that for individual speakers there is hardly any overlap, and that the STEEPNESS DIFFERENCE successfully categorises the two sibilants almost 100% of the time. None of the English or Dutch speakers nor the male Bengali speakers have any overlap at all, and the female speakers in Bengali overlap only marginally — i.e., $(\mathbf{a} - \mathbf{b})_{\int} > (\mathbf{a} - \mathbf{b})_{S}$ holds for individual subjects.

The frequency of STEEPNESS DIFFERENCE values for each language in absolute numbers and percentages is given in the middle part of Table I. The corresponding cumulative line graphs for individual speakers within each language are given in Fig. 5a–c. Based on the numbers shown in the middle part of Table I, we can now compute the optimal boundary value for each language. The optimal boundary value in db/kHz is defined to be the point where the classification errors for both consonants are minimal. The boundary values define right closed intervals, i.e., a boundary value of 10 means that all (**a–b**) values at and below 10 are classified as [s], and all values above 10 as [ʃ]. In Bengali, 8 is the optimal boundary point—only 1.4% of the [s]s and [ʃ]s are misclassified at this point. As for English, the boundary lies at 13, where only 0.7% of the [s]s do not fit. For Dutch, however, the pattern is not as straightforward. Unlike the other languages, it is difficult to find a satisfactory boundary point since the overlapping region is much larger: 10–15. The optimal boundary point is perhaps at 12, where 11.8% of the [s]s are misclassified, and 5.6% of the [ʃ]s do not fit. The main reason for Dutch being the most difficult to differentiate clearly is because DF2 places the boundary at 9, whereas the other three speakers differentiate at approximately 14. The range is therefore the same as that of the other languages, but there is a greater overlap between [s] and [ʃ]. This means that DF2 qualifies well within the Bengali categories. Thus, we observe that there are no clear language particular differences; rather, speakers place their boundary point within a given range. From Table I, we can see that the boundary values for individual English speakers vary just as much as those of the Dutch or Bengali speakers. In fact, had it not been for speaker EM1, the boundary value of English could have easily been at 7 and not at 13.

Computing across all languages, we see that although the maximal range of overlap could be 6–15, the boundaries at which languages seem to choose between [s] and [ʃ] lie between 8 and 13. Across all languages the optimal boundary point for the STEEPNESS DIFFERENCE measure would be 11, where 8.3% of the [s]s and 5.6% of the [ʃ]s are misclassified. These data are summarised in Table II, which gives the boundary values for the individual languages as well as for all the languages combined, together with the percentage of correct fit.

We next checked whether there was any influence of gender or the preceding vowel context. First, we computed the difference measure for each vowel type. Although the vowels of the three languages are not identical, it is possible to classify them broadly according to vowel height, backness, and roundness. Table III summarises the data for each vowel type, and Table IV gives the optimal boundary values together with the percentage of correct classification. The corresponding cumulative graphs are Fig. 6a

TABLE II. Optimal boundary values and percentages of correct classification for each language separately and across all data

	Boundary value	Percentage correct [s]	Percentage correct [ʃ]
Bengali	8	98.6	98.6
Dutch	12	88.2	94.4
English	13	99.3	100.0
All data	11	91.7	94.4

TABLE III. Overview of the data, grouped according to the vowel categories, high/mid/low, back/front, and round/unround. Since not all vowel categories contained the same number of words, percentages are given. See Table I for details on the presentation

(**a** − **b**)	≤ 6	7	8	9	10	11	12	13	14	15	≥ 16	∑
$High_s$%	75.0	6.9	2.8	2.1	3.5	3.5	2.8	2.8	0.0	0.7	.	100
$High_ʃ$%	.	.	0.7	0.7	2.1	5.6	5.6	6.9	2.8	5.6	70.1	100
Mid_s%	65.5	4.2	3.6	6.0	4.8	3.0	5.4	3.6	1.8	2.4	.	100
$Mid_ʃ$%	.	0.6	0.0	0.0	0.0	3.6	4.2	6.0	3.6	7.1	75.0	100
Low_s%	72.5	2.5	2.5	9.2	5.8	3.3	1.7	2.5	.	.	.	100
$Low_ʃ$%	.	.	.	0.8	0.8	1.7	2.5	4.2	6.7	3.3	80.0	100
$Back_s$%	74.2	4.2	2.1	6.3	3.3	3.3	2.9	2.5	0.4	0.8	.	100
$Back_ʃ$%	.	0.4	0.4	0.4	0.4	5.4	4.2	6.7	5.4	5.0	71.7	100
$Front_s$%	66.1	5.2	4.2	4.7	6.3	3.1	4.2	3.6	1.0	1.6	.	100
$Front_ʃ$%	.	.	.	0.5	1.6	1.6	4.2	4.7	2.6	6.3	78.6	100
$Unround_s$%	71.8	4.8	3.6	5.6	6.3	3.2	2.0	2.8	.	.	.	100
$Unround_ʃ$%	.	.	.	0.8	1.6	1.6	4.0	5.2	4.8	4.8	77.4	100
$Round_s$%	68.9	4.4	2.2	5.6	2.2	3.3	5.6	3.3	1.7	2.8	.	100
$Round_ʃ$%	.	0.6	0.6	0.0	0.0	6.7	4.4	6.7	3.3	6.7	71.1	100

TABLE IV. Optimal boundary values and percentages of correct classification for each vowel category

	Boundary value	Percentage correct [s]	Percentage correct [ʃ]
High	10	90.3	96.5
Mid	12	92.3	91.7
Low	11	95.8	96.7
Back	10	90.0	98.3
Front	12	93.8	92.2
Unrounded	11	95.2	96.0
Rounded	12	92.2	87.8

(height), b (backness), and c (roundness). The optimal boundary value for each vowel category lies between 10 and 12. As we can see from the three cumulative graphs, there is very little overlap between [s] and [ʃ] within each vowel category and the separation of the sibilants is independent of the vowel type. For instance, in Fig. 6b, the front and back

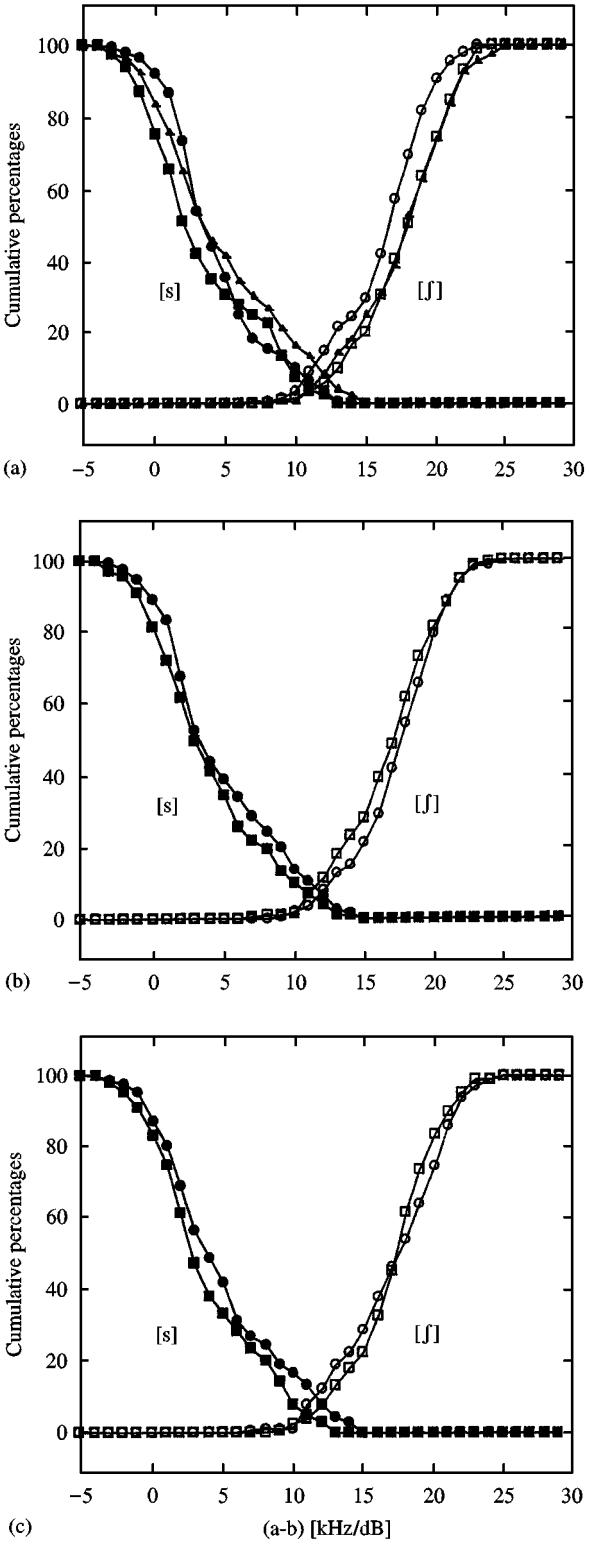

Cumulative percentages
100
80
60
40
20
0
[s]
[ʃ]
−5
0
5
10
15
20
25
30
(a)
(b)
(c)
(a-b) [kHz/dB]

TABLE V. Overview of the data, grouped according to gender. See Table I for details on the presentation

(**a** − **b**)	≤ 6	7	8	9	10	11	12	13	14	15	≥ 16	∑
$Female_S$	199	9	2	3	3	.	.	.	.	.	.	216
$Female_ʃ$	.	1	0	2	3	11	14	21	14	16	134	216
$Male_S$	106	11	11	21	17	14	15	13	3	5	.	216
$Male_ʃ$	.	.	1	0	1	5	4	4	4	4	189	216
$Female_S$%	92.1	4.2	0.9	1.4	1.4	.	.	.	.	.	.	100
$Female_ʃ$%	.	0.5	0.0	0.9	1.4	5.1	6.5	9.7	6.5	7.4	62.0	100
$Male_S$%	49.1	5.1	5.1	9.7	7.9	6.5	6.9	6.0	1.4	2.3		100
$Male_ʃ$%	.	.	0.5	0.0	0.5	2.3	1.9	1.9	1.9	3.7	87.5	100

TABLE VI. Optimal boundary values and percentages of correct classification for all male versus all female speakers

	Boundary value	Percentage correct [s]	Percentage correct [ʃ]
Female	9	98.6	98.6
Male	13	96.3	93.1

vowels show a nearly identical distribution. Not only do the lines for the front and back vowels stay close together, there is no systematic parallel shift of the lines — the lines for [s] cross — which would have indicated a front/back dependency. Similar patterns can be observed for height and roundedness. Thus, the results show that there is no real vowel influence on the discrimination of the sibilant fricatives.

The data are divided into gender categories in Tables V and VI. Fig. 7 is the corresponding cumulative line graph. The results indicate that the male and female speakers have different optimal boundary points: 13 dB/kHz for the male speakers and 9 dB/kHz for the female speakers. Looking at Fig. 7, one might argue that there is a gender dependency in the classification of [s] and [ʃ] because there is a parallel shift of the lines. However, although there is a difference, this does not indicate that the contrast is essentially different for men and women. Since most of the spectra are not situated immediately at the boundary, individual differences between speakers strongly influence the boundary point for a given category. As can be seen from Table I, the Bengali male speaker BM2 shows an optimal boundary value of 7, while EM1 differentiates the consonants at 15. On the other hand, a female speaker of Bengali, BF2, has a boundary value at 8, quite close to the male speaker BM2. In the cumulative graphs in Fig. 5, the

Figure 6. Cumulative percentages of STEEPNESS DIFFERENCE values for (a) all vowel height categories. Vowel height is indicated by means of symbols: circle = high, triangle = mid, square = low. (b) All front versus all back vowels. The symbols indicate backness of the vowel: circle = front, square = back. (c) All rounded versus all unrounded vowels. The symbols indicate roundedness of the vowel: circle = rounded, square = unrounded. Open and closed symbols: see legend for Fig. 4.

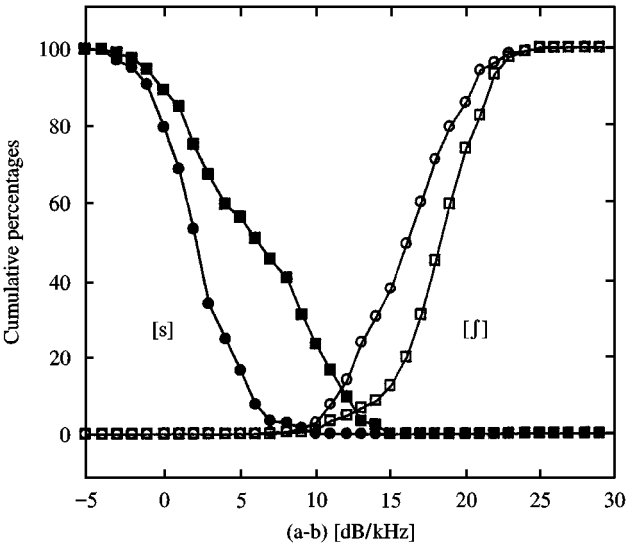


Figure 7. Cumulative percentages of STEEPNESS DIFFERENCE values for all female versus all male speakers. Speaker gender is indicated by means of symbols: circle = female, square = male. Open and closed symbols: see legend for Fig. 4.

parallel shifts of the lines indicate speaker variance. In Fig. 5b, there seems to be an apparent separation between males and females for [s], but not for [ʃ]. From the other graphs, it is obvious that there is no specific gender-based pattern. Although from Fig. 7, it seems that the male speakers always cluster to the right, in Fig. 5a, one female speaker is far to the right while the male speakers clusters together with the other female speaker. Thus, there is no gender dependency but rather a variation of individuals within a given range.

In sum, we have observed that the relationship $(\mathbf{a}-\mathbf{b})_{\int} > (\mathbf{a}-\mathbf{b})_{\mathrm{S}}$ holds for all speakers of the three languages we have investigated. The STEEPNESS DIFFERENCE can successfully categorise the two sibilants within a range of 7–15 dB/kHz across languages. There are, moreover, neither vowel nor gender dependencies. Rather, there is a certain amount of individual variation; speakers appear to place the boundaries within the given range. Further, this variation is independent of phonological status, i.e., the variance in English is just as much as in Dutch and Bengali. How this range relates to the theoretical points made with regard to the nature of universal phonetic features is discussed below.

4. Discussion

In this study, we investigated the claim that stable phonetic categories like the sibilants [s] and [ʃ] can be categorised acoustically by the same principle across languages, speakers, gender, and contexts, as well as phonological status. To that end, we quantified the observation that the frequency spectra of the two sibilants differ in the orientation of the spectral envelope within two frequency regions. We observed that the basic difference in the spectra for the two sibilants was the rate of increase in energy in the lower and higher frequencies. Choosing 2.5 kHz as the cutoff point between low and high

frequency, two regression lines, labelled **a** and **b**, were computed below and above this point for each spectrum. This rate of increase of energy was represented by the steepness of the slope of these regression lines. It was also noted that a single slope, **a** or **b**, could not distinguish the consonants, but that the difference (**a** − **b**) can successfully categorise the two sibilants. Computing a boundary point for each language, we found that the optimal boundary value lies between 8 and 13 for the three languages, although there is a range of STEEPNESS DIFFERENCE values 7–15 where there is a certain amount of overlap. Any variance in choosing the boundary is due to individual speaker variance and not dependent on language, vowel type, or gender.

As opposed to many recent papers (Bladon & Seitz, 1986; Forrest *et al*., 1988; Nittrouer *et al*., 1989; Nguyen *et al*., 1994), we deliberately refrained from using purely statistical measures to categorise the two sibilants. This choice was motivated by theoretical concerns. Although a statistical account may seem preferable because of the greater generality of the procedures involved, this advantage is only apparent. We believe that an analysis of speech sounds should not be separated from theories of speech production and perception. A greater statistical generality does not necessarily add to our understanding of the acoustic correlates of features. Clearly a feature cannot be reduced to the outcome of a statistical procedure, but should ideally be related to specific acoustic and articulatory correlates. Neither does the human perceptual system work like a general statistical device but is adapted to its specific task of extracting the relevant properties of the signal: amplitudinal, spectral, and durational properties, including their change over time. Therefore, we based our analysis on a quantification of the observed spectral shape which can be directly correlated with the articulatory configuration required to produce these consonants, and which can give the basis for perceptual discrimination between them.

Although our analysis uses the concept of linear regression, this statistical measure only quantifies the observed differences in steepness of the spectral slopes. Such a procedure is in principle quite different from treating spectra as input to a statistical procedure, trusting that a value may be obtained to distinguish between categories. For instance, Forrest *et al*. (1988) calculated four statistical moments, but did not use the first two because they turned out to be of no help in distinguishing the consonants. There was no particular reason for choosing this procedure; it was not based on any language production or perception concerns, and many other procedures could have been taken instead. Note that exactly the opposite was done in the investigation by Nguyen *et al*. (1994), where the outcome of a statistical principal component analysis, in particular the first two factors, led to the conclusion that the spectral area between 2 and 4 kHz is important for a differentiation between [s] and [ʃ]. The third factor, which is also used in their analysis, seems to quantify the energy distribution "in more specific parts of the spectrum" (Nguyen *et al*. 1994: 35). It remains unclear, however, which specific parts are meant, and why they play a role in differentiating the two sibilants. Actually, a correct classification of about 90% across all categories in our metric shows that a speech production oriented approach can do as well — if not better — than a statistical one.

We computed the skewness and kurtosis as described by Forrest *et al*. (1988: 117–118) for our data set with a linear frequency scale. The skewness values for 80% of all [ʃ] lay between −0.50 and 2.10, and of 80% of all [s] lay between −1.23 and 1.14. The kurtosis values for 80% of all [ʃ] lay between −1.78 and 5.43, and of 80% of all [s] lay between −1.20 and 4.29. This means that the skewness values of both sibilants show a considerable overlap and that the kurtosis values for [s] are nearly entirely included in the

values for [ʃ]. For example, choosing a kurtosis value of 1 as upper limit for [ʃ] catches 75% of all [ʃ] but includes 78% of all [s] as well. If the kurtosis value is higher than 3, then only 7% of all [s] are included, but unfortunately also only 14% of all [ʃ] fulfill this criterion. In other words, skewness and kurtosis have only a very limited discriminating power to separate the sibilants across all speakers.

A further point regarding our methodology is that we based our analysis on FFT spectra and not on Bark spectra. Most of the statistical methods use a Bark-scaled spectrum as the basis of their analysis (Bladon & Seitz, 1986; Forrest *et al*., 1988; Nittrouer *et al*., 1989; Nguyen *et al*., 1994), because the Bark scale has been stated to be an appropriate spectral representation of the perceptual system. As already mentioned earlier, we do not consider Bark scaled spectra — or any other similar spectral transformation — to be a suitable measure for fricative noise. In fact, Rosner and Pickering (1994), giving a brief overview of various scales (log, mel, Koenig, Bark, and ERB-rate), conclude that the ERB-rate scale is better motivated than the Bark scale. Moreover, all these scales are meant to optimise the perceptual vowel space. Bark spectra represent the perception of frequencies in terms of critical bands, expressing the notion that adjacent frequency peaks in a spectrum are perceived as one peak if they appear in the same critical band (Fletcher, 1940). For an overall distributional measure describing the gross spectral shape, it is not this differentiating quality of our auditory system that is relevant, but rather the global representation of the spectral shape, as represented in a power spectrum. In fact, the decrease of discrimination performance by Forrest *et al.* (1988) for the female speakers indicates that the Bark scale did not invariably improve the categorization.

The Bark scale (Zwicker, 1961) and the ERB scale (Equivalent Rectangular Bandwidth, Patterson, 1976) differ in two ways from the linear frequency scale that we use in our calculations. First, they are non-linear in the sense that the lower frequency range is stretched, while the higher frequency range is compressed. Roughly speaking, the scales are more linear for lower frequencies (approximately below 1000 Hz) and more logarithmic for higher frequencies. This reflects the auditory system, which resolves lower frequencies better than higher frequencies. Second, these scales perform a non-monotonic modification of the spectrum, since spectral energy within a critical band contributes to one point on the transformed scale. This causes spectral peaks that are close together to be mapped onto one spectral peak, which again reflects a feature of the auditory system. It is not always the case that in using the Bark and ERB scale both properties are exploited. For instance, Syrdal and Gopal (1986) performed only the non-linear transformation of formant frequencies, whereas Forrest *et al.* (1988) used both properties in their computation of their spectra.

However, using a Bark or ERB scaled spectrum does not have a greater impact on the separability of spectral peaks in one critical band for sibilants. Spectra of friction are smooth (Stevens, 1989: 23; Fant, 1960: 26) and are better classified by the overall spectral shape than by the positions of formant peaks. Therefore, separating or collapsing spectral peaks with critical bands is not an issue, and using Bark or ERB transformations is a rather complicated way of performing spectral smoothing for sibilants. Further, the non-linearity of these scales will simple change the shape of the regression line in our analysis. Instead, we think that the spectral shape is well approximated by linear curves on a linear scale. Using more complicated regression lines – or, what is equivalent, using a non-linear transformation prior to the computation of linear regression lines — is not necessary to capture spectral shapes of sibilants.

Are the results of our investigation compatible within the notion of acoustic invariance? Past research has convincingly suggested that this invariance is to be conceived as relative rather than absolute (Blumstein & Stevens, 1981; Kewley-Port, 1983; Lahiri *et al.* 1984). *Relative* can either be understood as referring to the relation between different spectra in time (Kewley-Port, 1983; Lahiri *et al*., 1984) or to the relation within spectra (Jongman *et al*., 1985), as done in this study. In either case, a relational nature of a metric can normalise across speakers, gender, contexts, and other sources of variation. We have seen that the boundary value between [+ anterior] and [− anterior] sibilant categories is slightly different in the three languages we examined, but that the acoustic dimension is essentially the same in all these languages.

Our analysis using the STEEPNESS DIFFERENCE measure showed a clear categorization overall, but also some variability in the location of the boundary between the two categories. It turned out that this variability could not be correlated with one particular categorization of the data. Neither language, gender, nor vowel context crucially affects the acoustic distinction. Rather, any observed variance appears to be part of speaker variation, most clearly seen in Dutch. For instance, the Dutch speaker DF2 had a boundary value of 8 dB/kHz rather than the general language value, which seems to be at 12 dB/kHz.

Another interesting outcome of the research is that there is no systematic difference between languages that have a strict phonemic difference between the two sibilants like English as against Bengali and Dutch, which have an allophonic contrast. Recall that we chose to study these three languages to not only find a measure to differentiate the two sound categories effectively, but also to test whether there is a difference between languages with a phonemic contrast and languages with an allophonic contrast. One might predict that a phonemic difference implies a more categorical differentiation, while a phonetic difference may allow a more lax distinction, i.e., a gradient along some acoustic dimension. In contrast to this prediction we found no difference in the realisation of the two sounds under different phonological conditions — i.e., phonemic status did not have any effect on discriminability. Regardless of the difference in the phonological status of the feature [anterior] in the three languages, the two segments were clearly separated in the individual languages within a fixed range of values (see Table I). This outcome should not be confused with the results of Utman and Blumstein (1994), who suggested "that while the fundamental manifestation of the acoustic property is the same across languages, its instantiation may be influenced by the functional role that its associated feature plays in the language" (p. 221). While comparing the labiodental fricative [f] in Ewe and English, they found that the English [f] was acoustically less distinct from the Ewe bilabial fricative [ɸ] than the Ewe [f]. The difference in the distinctiveness in their investigation must be attributed to the lack of one of the sounds in English, while in our research both sounds occur in all three languages but differ in their phonological status. As a result, the sibilant sounds in Dutch and Bengali, where only an allophonic contrast exists, are not less discriminable than in English, which has a phonemic opposition.

In this study, the sounds we investigated differ in their place of articulation and the difference is quantal in nature (Stevens, 1989). Contrasts like [f] and [ɸ], both of which are labial, are not quantal, and perhaps the difference between these sounds would be less discriminable if they were allophonic in a language. Similarly, other phonological features like voicing or aspiration, or rounding in vowels may be less discriminable when they contrast allophonically. However, this remains to be seen.

To conclude, the contrast between the sibilants [+ anterior] [s] and [− anterior] [ʃ] can effectively be expressed with a simple gross spectral measure of only 40 ms in the fricative noise, relating the spectral slope below and above 2.5 kHz into one numerical value. The categorization it provides is robust, allowing a correct classification of 90% of the fricatives across vowel context, gender, and language categories. The steepness metric is a distributional measure, quantifying gross spectral shape rather than absolute properties of the signal or the spectrum, like frequency peaks or locus of formant transitions. As such, it is compatible with recent accounts of acoustic invariance which view invariance as being relative rather than absolute. The present analysis specifies a dimension along which languages may situate a particular phonetic distinction. Finally, the results suggest that an acoustic distinction like anteriority in sibilants is realised in the same way in different languages irrespective of whether the contrast in question is phonemic or allophonic.

We would like to thank the editors Terrance Nearey and Bruce Derwing for their suggestions and the reviewers Björn Lindblom, Ken Stevens, and Gary Weismer for their valuable comments. This research was partially funded by a DFG grant to the third author.

References

Behrens, S. J. & Blumstein, S. E. (1988) Acoustic characteristics of English voiceless fricatives, *Journal of Phonetics*, **16,** 295–298

Bladon, A. & Seitz, F. (1986) Spectral edge orientation as a discriminator of fricatives, *Journal of the Acoustical Society of America*, Suppl. 1, **80,** S18–S19

Blumstein, S. E. & Stevens, K. N. (1981) Phonetic features and acoustic invariance in speech, *Cognition* **10,** 25–32

Booij, G. (1995) *The phonology of Dutch*. Oxford: Clarendon Press

Chatterji, S. K. (1975) *The origin and development of the Bengali language*. Calcutta: Rupa & Co

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. New York: Harper & Row

Clark, C. & Bladon, A. (1986) Analysis and perception of sibilant fricatives: Shona data, *The Journal of the Acoustical Society of America*, Suppl. 1, **80,** S125

Collins, B. & Mees, I. (1984) *The sounds of English and Dutch*. Leiden: Leiden University Press

Fant, G. (1960) *Acoustic theory of speech production*. The Hague: Mouton

Fletcher, H. (1940) Auditory patterns, *Reviews of modern physics*, **12,** 47–65

Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988) Statistical analysis of word-initial voiceless obstruents: preliminary data, *Journal of the Acoustical Society of America*, **84,** 115–123

Gussenhoven, C. & Broeders, A. (1976). *The Pronunication of English*. Groningen: Wolters-Noordhoff-Longman

Halle, M. & Clements, G. N. (1983) *Problem book in phonology*. Cambridge, MA: MIT Press

Heinz, J. M. & Stevens, K. N. (1961) On the properties of voiceless fricative consonants, *Journal of the Acoustical Society of America*, **33,** 589–596

Hughes, G. W. & Halle, M. (1956) Spectral properties of fricative consonants, *Journal of the Acoustical Society of America*, **28,** 303–310

Jakobson, R., Fant, G. & Halle, M. (1952) *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, MA: MIT Press

Jongman, A. (1989) Duration of frication noise required for identification of English fricatives, *Journal of the Acoustical Society of America*, **85,** 1718–1725

Jongman, A., Blumstein, S. E. & Lahiri, A. (1985) Acoustic properties for dental and alveolar stop consonants: a cross-language study, *Journal of Phonetics*, **13,** 235–251

Keating, P. (1991) Coronal places of articulation. In *The special status of coronals* (C. Paradis & J.-F. Prunet, editors), pp. 29–48. London: Academic Press

Kewley-Port, D. (1983) Time-varying features as correlates of place of articulation in stop consonants, *Journal of the Acoustical Society of America*, **73,** 322–335

Ladefoged, P. (1982) *A course in phonetics*. San Diego: Harcourt Brace Jovanovich

Lahiri, A. (1991) Anteriority in sibilants. In *Proceedings of the XIIth International Congress of Phonetic Sciences*, pp. 384–388

Lahiri, A. & Evers, V. (1991) Palatalization and coronality. In *The special status of coronals* (C. Paradis & J.-F. Prunet, editors), pp. 79–100. London: Academic Press
Lahiri, A., Gewirth, L. & Blumstein, S. E. (1984) A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study, *Journal of the Acoustical Society of America*, **76,** 391–404
Lindblom, B. (1983) Economy of speech gestures. In *The production of speech* (P. F. MacNeilage, editor), pp. 217–245. New York: Springer
Lisker, L. & Abramson, A. S. (1964) A cross-language study of voicing in initial stops: acoustical measurements, *Word*, **20,** 384–422
Maddieson, I. (1984) *Patterns of sounds*. Cambridge: Cambridge University Press
Manuel, S. Y. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *Journal of the Acoustical Society of America*, **88,** 1286–1298
McCarthy, J. J. (1988) Feature geometry and dependency: a review, *Phonetica*, **43,** 84–108
Nguyen, N., Hoole, P. & Marchal, A. (1994) Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters, *Journal of the Acoustical Society of America*, **96,** 33–39
Nittrouer, S., Studdert-Kennedy, M. & McGowan, R. S. (1989) The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults, *Journal of Speech and Hearing Research*, **32,** 120–132
Paradis, C. & Prunet, J.-F. (1991) Introduction: asymmetry and visibility in consonant articulations. In *The special status of coronals* (C. Paradis & J.-F. Prunet, editors), pp. 1–28. London: Academic Press
Patterson, R. D. (1976) Auditory filter shapes derived with noise stimuli, *Journal of the Acoustical Society of America*, **59,** 640–654
Rosner, B. S. & Pickering, J. B. (1994) *Vowel perception and production*. Oxford: Oxford University Press
Stevens, K. N. (1971) Airflow and turbulence noise for fricative and stop consonants: static considerations, *Journal of the Acoustical Society of America*, **50,** 1180–1192
Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics*, **17,** 3–45
Strevens, P. (1960) Spectra of fricative noise in human speech, *Language and Speech*, **3,** 32–49
Syrdal, A. K. & Gopal, H. S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels, *Journal of the Acoustical Society of America*, **79,** 1086–1100
Utman, J. A. & Blumstein, S. E. (1994) The influence of languages on the acoustic properties of phonetic features: A study of the feature [strident] in Ewe and English, *Phonetica*, **51,** 221–238
Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands (Frequenzgruppen), *Journal of the Acoustical Society of America*, **33,** 248

Appendix

List of words used in the investigation. Bengali forms are in phonetic transcription, Dutch and English in standard orthography

[ʃ]		[s]		Vowel	High	Back	Low	Round
BENGALI								
miʃʈi	'sweet'	kʰisʈi	'abuse'	[i]	+	−	−	−
piʃʈʰo	'crushed'	pis̪ʈol	'pistol'					
biʃʈi	'rain'	dis̪ʈe	'quire'					
duʃʈu	'naughty'	dus̪ʈʰo	'poor'	[u]	+	+	−	+
kuʃʈi	'horoscope'	kus̪ʈi	'wrestling'					
puʃʈi	'nourishment'	pus̪ʈɔk	'book'					
kɔʃʈo	'suffering'	hɔs̪ʈo	'hand'	[ɔ]	−	+	−	+
nɔʃʈo	'waste'	mɔs̪ʈo	'large'					
spɔʃʈo	'clear'	bɔs̪ʈa	'sack'					
kaʃʈʰo	'wood'	kʰɑs̪ʈa	'crisp'	[ɑ]	−	+	+	−
bɑʃpo	'vapour'	bɑs̪ʈob	'reality'					
mɑʃʈar	'tutor'	rɑs̪ʈa	'road'					
DUTCH								
misje	'little mass'	missen	'to fail'	[ɪ]	+	−	−	−
nisje	'little niche'	nissen	'niches'					
visje	'small fish'	vissen	'fishes'					

Appendix (Continued).

busje	'small bus'	bussen	'buses'	[œ]	−	−	−	+
musje	'bird'	mussen	'birds'					
zusje	'little sister'	zussen	'sisters'					
bosje	'small forrest'	bossen	'forests'	[ɔ]	−	+	−	+
osje	'small ox'	ossen	'oxen'					
vosje	'small fox'	vossen	'foxes'					
pasje	'credit card'	passen	'credit cards'	[ɑ]	−	+	+	−
tasje	'small bag'	tassen	'bags'					
wasje	'small wash'	wassen	'to wash'					
AMERICAN ENGLISH								
fishing		kissing		[ɪ]	+	−	−	−
mission		missing						
pushy		pussy		[ʊ]	+	+	−	+
session		lesson		[ɛ]	−	−	−	−
pressure		pressing						
special		vessel						
flushing		fussing		[ʌ]	−	+	−	−
hushing		hussy						
squashing		fossil		[ɒ]	−	+	+	+
washing		bossy						
bashing		basset		[æ]	−	−	+	−
lashing		lassie						