Evaluating a new measure of fricative source intensity

Matthew Faytak and Keith Johnson (University of California, Berkeley)

Background. Fricative spectral shape is a common focus of acoustic analysis, as reflected by the most commonly used acoustic measures for describing fricatives: the first four spectral moments, location of peak frequency, and more complex measures approximating spectral slope (e.g. Maniwa et al, 2009). A separate question asked less commonly concerns the amplitude of the fricative noise source rather than of the received spectrum. In many applications it is useful to be able to measure source amplitude separately from properties of the radiated sound wave, for instance in studies of fricative aerodynamics (Solé, 2010) theoretical acoustics (Stevens, 1998), fricative acoustic and articularatory dynamics (Iskarous, et al., 2011), and unusual segmental contrasts (Ling, 2007). However, existing means of extracting source amplitude are either highly complex (pitch-scaled decomposition, Jackson and Shadle 2000) or lack broad applicability (harmonic-to-noise ratio, which cannot be applied to voiceless phones, see use in e.g. Ling, 2007; Maniwa et al, 2009).

The measure. We put forward the Sonorant-Fricative Discriminant Index (SFDI), described in Ananthapadmanabha et al. (2014), as a suitable and computationally inexpensive approximation of fricative noise source amplitude. The SFDI is calculated as the arctan of the sum of the LPC coefficients measured on a pre-emphasized waveform window. Ananthapadmanabha et al. (2014) point out that the DC gain of the digital inverse filter $A(z)$ at $z = 1$ (or frequency = 0) is given by $A(z=1) = 1 + a_1 + a_2 + a_3 + \ldots + a_M$; this expression corresponds to the sum of the LPC coefficients. Since linear prediction models an all-pole filter and the periodic voice source energy is returned in the LPC residual, it is suggested that fricative noise sources are captured in the DC gain of the LPC filter.

Validation. We calculated the SFDI for seven time points (start, end, and five points evenly spaced in between) in all automatically aligned (Yuan and Liberman, 2008) non-silence segments in readings of a 100-word list by 34 male native speakers of American English. Pre-emphasized audio ($p = 0.96$) was downsampled to a 16 kHz sampling rate. A window of 0.02 s and an LPC order of 18 was used for calculation of the SFDI. Results illustrate that most phones in the broad class of fricatives (including /h/) have considerably higher midpoint SFDI when compared to sonorants and vowels (Fig. 1, Fig. 2). Canonical strident fricatives /s/, /ʃ/, /f/ have by far the highest SFDI values. The phones /ð/, /v/, /h/ behave exceptionally, with markedly low SFDI values resembling approximants (e.g. /w/, /j/). The lower SFDI observed for voiced /z/, /ʒ/ relative to their voiceless counterparts is expected given their weaker noise source, due to a lower pressure gradient across the fricative constriction (Stevens, 1998:100-109). A rising-falling “arc” is observed in fricatives to the exclusion of other phone classes (Fig. 3); this peaking pattern accords with other observed fricative spectral dynamics (Iskarous et al, 2011).
Figure 1: Untransformed SFDI as a function of measured RMS amplitude for all segment midpoints. Regression lines are plotted to fit the data for fricative and non-fricative segment classes. Note that observed fricative SFDI values tend to be higher than non-fricatives’ at all levels of RMS amplitude.

![SFDI at midpoint, all continuants](image1)

Figure 2: SFDI of phone midpoints; outliers excluded. Non-continuants (stops and affricates) and [ə] are omitted. Phones are labeled with their ARPABET transcriptions.

![SFDI time series, continuants by class](image2)

Figure 3: SFDI trajectory pooled by segment class for timepoints (‘idx’) 2-6; starting and ending timepoints are excluded due to considerably higher variability. The phones /v/, /ð/, and /h/ are included in the “fricative” class and contribute to the high variance observed.

References


