TOWARDS THE SPECTRAL CHARACTERISTICS
OF FRICATIVE CONSONANTS

Christine H. Shadle1, Pierre Badin2, André Moulinier1

University of Southampton, UK1
ICP, INPG, Grenoble, France2

ABSTRACT
Articulatory data and the all-pole transfer functions for sustained fricative
consonts [s, f, θ] were used to identify the cavity affiliation of peaks
and troughs in the far-field spectra. This identification then allowed an
analysis of the differences between fricatives, and across subjects within
fricatives, necessary steps towards the establishment of distinguishing acous-
tic cues. Finally we use this articulatory-acoustic mapping to explain some of
the across-subject differences. This sequence should lead to a set of paired
articulatory-acoustic cues that can then be tested for their perceptual
importance.

1. INTRODUCTION
It has long been established that [s, θ] are distinguished chiefly by the transi-
tions of the vowels on either side, while [s, f] are distinguished by their
spectral characteristics [3]. But establishing the particular spectral cues that
distinguish [s, θ] from each other, or from other fricatives, is more difficult.
Many authors report consistency within a speaker, but high variability
across speakers [4,7]. Perhaps as a result, efforts to phrase distinguishing
cues in terms of the frequency range of the highest intensity levels, or in terms
of relative intensity levels, seem to work well within a speaker but poorly
across speakers (e.g. the frequency ranges overlap so much as to be use-
less.) [7].
In this study we explore variability in the spectra of sustained fricatives.
First, we need to establish which aspects of the spectrum are consistent
within a speaker-fricative combination. Then where possible we identify the
articulatory parameters that control these consistent features of the spec-
trum. Finally we use this articulatory-acoustic mapping to explain some of
the across-subject differences. This sequence should lead to a set of paired
articulatory-acoustic cues that can then be tested for their perceptual
importance.

2. METHOD
2.1 Corpus and Speakers
The corpus used in this paper is the result of a larger study (Leeds, Gren-
oble, Southampton). It includes articulatory, aerodynamic and acoustic
measurements made of two speakers.
The corpus includes 13 fricatives [s, θ, θ, f, θ, g, θ, θ, θ, θ, θ, h] produced in several
ways. This study refers only to the sus-
tained corpus, in which the set of 13
fricatives was said six times; in each
set, the order of the 13 fricatives was
randomized. Two different recordings of the sustained fricatives were used
in this study, as detailed below.
The two speakers used for the corpus
are the first two authors of this paper,
and will be referred to as CS, a woman
speaker of General American English,
and PB, a man speaker of French.
Although the list of fricatives recorded
includes several that are not native to
either speaker, these were included
deliberately to obtain further examples
of place variation for the same vocal
tracts.
In addition to measurements made
while speaking, X-ray data and dental
impressions were available for each
subject. Together with EPG data and
external photographs, these were used
to construct an area function for each
unvoiced fricative for each speaker [6].

2.2 Acoustic Analysis
Data shown in this paper were recorded
under high-fidelity conditions: the
subject was seated in a chamber
anechoic above 170 Hz, with a B&K
4165 ¼" microphone located 1 m
in front of the subject's mouth. Record-
ings were made with a Sony PCM
system at 16 bits with a sampling
frequency of 44.1 kHz. A calibration
signal was recorded to allow absolute
sound pressure level to be retained.
An average power spectral density
function was computed by averaging
25 spectra in the center of the 3 s frica-
tive. Each spectrum was computed
using a 20m Hanning-window.

2.3 Determination of Transfer Function
In this experiment, the subject
assumed the position for a fricative,
but without actual speech production
(glottis held closed). The vocal tract
was excited by a small loudspeaker fed
with white noise and pressed against
the neck just above the thyroid carti-
lage. A microphone located 2 cm from
the mouth detected the (very weak)
noise signal after filtering by the vocal
tract. This signal was essentially the
all-pole transfer function of the tract,
up to about 5 kHz.
The area functions derived from articulatory
data were then used to predict the all-pole transfer function
for each fricative. Comparison of pre-
diction and measured all-pole functions
then enabled identification of the cav-
ity affiliation of each pole. Further
details are given in [2].

3. RESULTS AND DISCUSSION
Figure 1 shows three of the fricatives
analyzed, with all six tokens shown on
each graph. Note first the consistency
apparent within each graph, i.e. within
each fricative-subject combination.
This consistency makes it easier to
evaluate the variability across
speakers, and across fricatives. For
[s, θ] the overall spectral shapes are
similar but the frequencies at which
particular peaks occur differ between
the two speakers. For [θ], even the
overall shape differs: both speakers
have a region of high energy between
1.5 and 6 kHz for PB, and 2.5 to 7
kHz for CS. However, for PB there is
an abrupt drop in amplitude of some
10 dB at 6 kHz and the spectrum is
approximately level above that fre-
quency; for CS, there is no abrupt
drop. Instead the level falls off stead-
ily, decreasing 20 dB between 7 and
12 kHz. Can we make sense of these
differences?
Badin's results [2] indicate that for
CS's [θ], F1 is a Helmholtz resonance of
back cavity and constriction; F2 and
F3 are back cavity resonances; and F4
is a front cavity resonance. A series
pressure source in the front cavity
would result in zeros cancelling the
back cavity resonances, plus two free
zeros: one corresponds to a Helmholtz
resonance of the constriction and the
part of the front cavity between the
constriction exit and the source. The
other corresponds to the half-
wave-length resonance of the same part
of the front cavity.

Since CS has smaller vocal tract dimen-
sions, her formant frequencies are
predicted to be higher, and in fact they
are. However, a much more obvious
difference is that for CS the first four
formants are approximately evenly
spaced, while PB has F2, F3 and F4
clustered together. With the zeros
terminated, these small differences in
formant frequency make a big differ-
ence in formant amplitude: for PB, the
second formant is boosted and becomes
the lowest high-amplitude peak, while
for CS, F3 takes on that role. This
means that the lower edge of the high-
amplitude region differs by 1 kHz, even
though F2 differs by only 100-200 Hz.

Above 5 kHz we have less informa-
tion to work with. However, the dif-
ferences in spectral amplitude and
slope could be explained if the free
zero were at a significantly lower fre-
quency for PB than for CS, e.g. 7 and
12 kHz respectively. This zero fre-
quency should be inversely propor-
tional to ip, the teeth-constriction
distance, and in fact ip is significantly
longer for PB, as evidenced from X-ray and direct
palatography. This is surprising since
the vocal tract dimensions in the anterior part of the mouth cavity, obtained from measurements of the two subjects' dental impressions, are quite similar. Since the phoneme is native to each subject, and the spectral differences noted are consistent within each subject, more subjects are needed to establish why the articulatory differences exist.

The fricative [s] is more similar for the two subjects. Since the front cavity is smaller than for [f], the corresponding resonances are higher. For CS, it appears from transfer function simulations that the lowest front cavity resonance is F6 (see Fig. 1); F2, F3, F4 and F5 are the lowest resonances of the back cavity (harmonics of the half-wavelength mode), and are accompanied by bound zeros. For PB the lowest front cavity resonance is F5, and F2, F3, and F4 are the back cavity resonances [1]. The differences between these resonances are consistent with the articulatory data. The amplitude of the plateau above the front cavity resonance relative to the spectral level of this resonance varies noticeably between the two subjects, and again the free zero may be lower for PB (approximately 11.5 kHz) than for CS (well above 12 kHz).

The fricative [c] is not native to either speaker, and so might be expected to be more variable. In fact, it looks consistent for each speaker, and the overall spectral shape is similar. For both speakers, the lowest front-cavity resonance is the lowest high-amplitude formant. This corresponds to F4 for CS, F3 for PB. Although the front cavity is longer for [c] than for [f], this front-cavity resonance is not significantly lower. A possible explanation is that for extremely short front cavities, the resonance frequency is related to the volume or possibly vertical dimension. Thus the exact shape of the sublingual cavity becomes important for [s,f]. As for [f], the spectral shape at high frequencies differs, and could be explained in part by a difference in source-constriction distance. The likelihood that the source is distributed

5. CONCLUSION
The search for acoustic cues distinguishing fricative consonants must begin with a study of the variability present in fricative production. By using subjects for whom much articulatory data is available, it has been possible to locate low-amplitude but consistent spectral peaks, and to discover their cavity affiliation and controlling parameters. Although vocal tract dimensions influence peak frequencies, the added complications introduced by zeros mean that simple measures such as frequency range for high-amplitude regions are likely to be highly variable.

6. ACKNOWLEDGEMENT
This work was funded in part by a collaborative EC SCIENCE award, CECSCI*0147C(EDB).

7. REFERENCES

Fig. 1. Averaged power spectral densities for sustained productions of the fricatives [s,f,c]. In each graph the six curves shown correspond to the six tokens uttered by each subject. Subject PB is male, native French speaker; CS is female, native General American speaker.