PARAMETERIZING SPECTRAL CHARACTERISTICS OF EUROPEAN PORTUGUESE FRICATIVES

Luis M. T. Jesus and Christine H. Shadle

Department of Electronics and Computer Science
University of Southampton, Highfield, Southampton, SO17 1BJ, UK

e-mail: {lmtj97r, chs}@ecs.soton.ac.uk

www: http://www.ecs.soton.ac.uk/~{lmtj97r, chs}

INTRODUCTION

As part of our on-going study of Portuguese fricatives, we have developed corpora that include real words and phonologically-possible nonsense words, and performed acoustic analysis of recordings of four native speakers.

We seek a way of parameterizing the fricatives that makes use of our knowledge of the underlying aeroacoustics.

Findings so far: **devoicing** in over one half of voiced fricatives, especially in destressed syllables; a correspondence between higher effort level and **syllable stress**; and some effect of **vowel context** on the fricative spectrum (Jesus and Shadle, 1999).

Multiple comparisons possible in such a study:

- across speakers,
- corpora,
- place,
- vowel context,
- syllable stress,
- location within fricative, etc.

Parameterization essential to handle large corpus.

PREVIOUS STUDIES PARAMETERIZING FRICATIVES

Two goals: distinguish fricatives (e.g. speech recognition); characterize fricatives (e.g., voiced-voiceless, normal-disordered speech).

Three approaches: locus equations; spectral moments; other time- or frequency-domain parameters.

- **Wilde (1995)**
  - Goals: characterize natural fricatives and assess perceptual importance of parameters via synthesis.
  - Approach: various time/frequency-domain parameters.
  - Corpus: 4 speakers of American English. Nonsense words.
  - Results: voiceless fricatives are more dependent on vowel context, and voicing onset time and formant structure provide important place information. Amplitudes of fricative noise in restricted frequency regions can distinguish sibilants from nonsibilants.

- **Jongman et al. (1995, 1998)**
  - Goals: distinguish fricatives.
  - Approach: spectral moments, locus equations, and other time/frequency-domain parameters.
  - Corpus: 3 speakers of American English (Jongman et al., 1995) and 20 speakers of American English (Jongman et al., 1998). Nonsense words.
  - Results: spectral peak location and noise duration distinguished sibilants from non-sibilants; spectral peak location separated /s,z/ from /ʃ,ʒ/; the amplitude distinguished all four places of articulation. The slope of locus equations could be used to differentiate labiodental from the other three places of articulation. The first moment distinguished all places of articulation.
- **Sussman (1994)**
  - Goals: distinguish fricatives.
  - Approach: locus equations.
  - Corpus: 4 speakers of American English. Small number of nonsense words.
  - Results: although this technique worked well on stops, it wasn't very successful on fricatives.

- **Evers et al. (1998)**
  - Goals: distinguish and characterize fricatives.
  - Approach: various time/frequency-domain parameters.
  - Corpus: two speakers of American English, Bengali and Dutch. 12 real words.
  - Results: it was possible to separate /s/ from /ʃ/ by using the difference in slope below and above 2.5 kHz.
- **Forrest et al. (1988)**
  - Goals: distinguish and characterize fricatives.
  - Approach: spectral moments.
  - Corpus: 10 speakers of American English. 5 real words.
  - Results: worked well to classify stops but could not distinguish all fricatives.
- **Shadle and Mair (1996)**
  - Goals: distinguish and characterize fricatives.
  - Approach: spectral moments and two additional spectral parameters (dynamic amplitude A_d and spectral slope S_p).
  - Corpus: one American English speaker, one French speaker. Large number of sustained fricatives and nonsense words.
  - Results: the spectral moments proved not to be useful for distinguishing fricatives when used on multiple tokens, varying effort levels, different vowel contexts, and three different locations within a fricative. A_d and S_p did not distinguish the fricatives completely but did vary with source location and effort level as predicted.

METHOD

Corpus of Portuguese fricatives /f, v, s, z, ʃ, ʒ/:

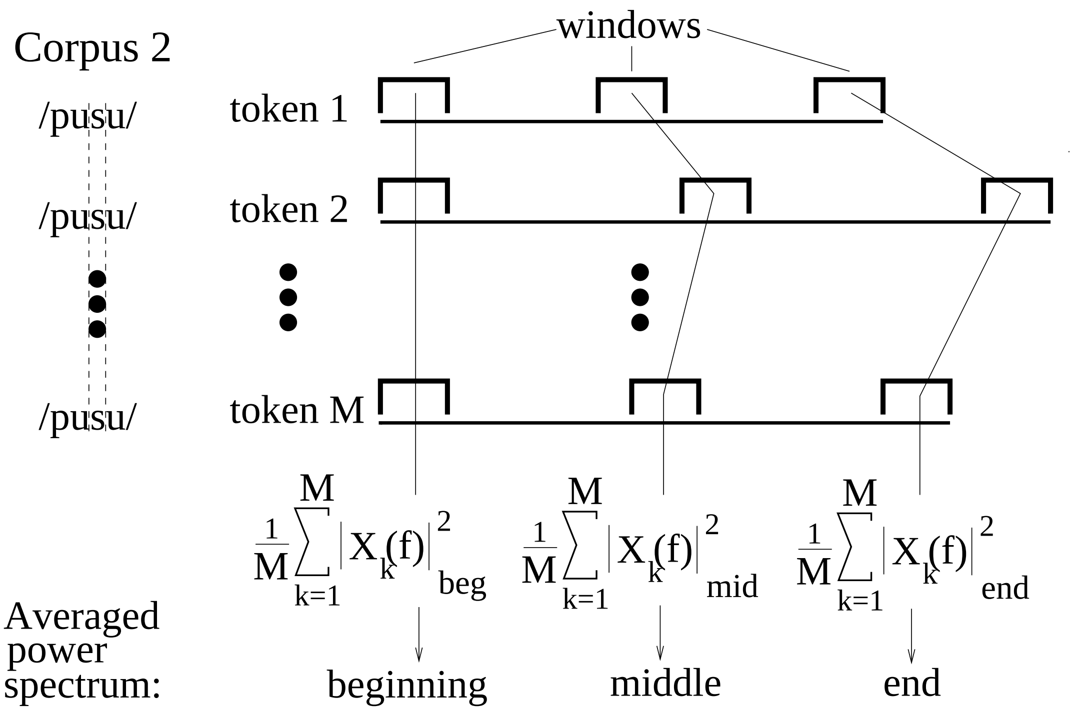
- **Corpus 1a** – sustained fricatives preceded by vowels /ɨ, ɐ, u/;
- **Corpus 1b** – fricatives sustained at different effort levels;
- **Corpus 2** – /pV_1FV_2/ repeated 12 times on one breath, V_i =/ɨ, ɐ, u/, and following Portuguese phonological rules;
- **Corpus 3** – 154 real words produced in a frame sentence. Examples of all Portuguese fricatives, with all non-nasalized vowels.

Subjects: 4 adult Portuguese native speakers; two males (**LMTJ** and **CFGA**) and two females (**ACC** and **ISSS**).

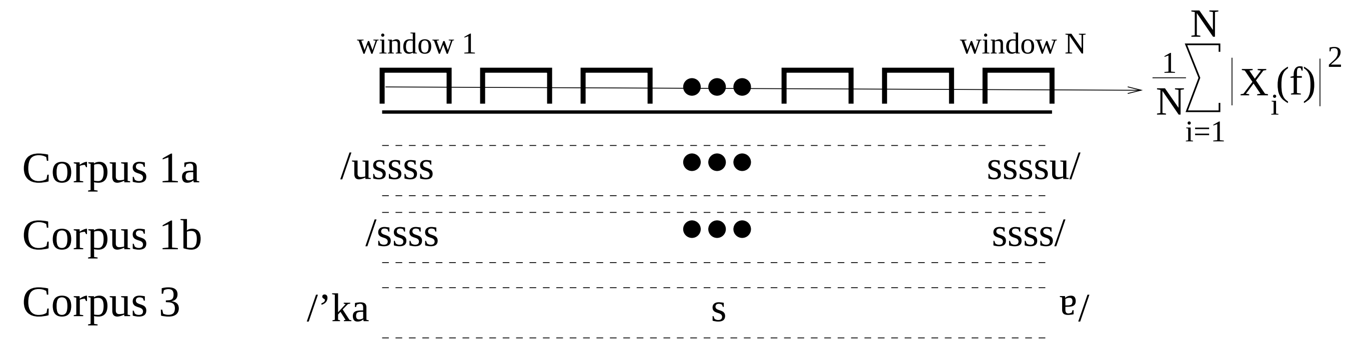
Recordings: sound-treated booth (B & K 4165 1/2 inch microphone 1 m from the subject, B & K 2636 measurement amplifier), Sony TCD-D7 DAT recorder (16 bits, sampling frequency 48 kHz).

Analysis: averaged power spectra were computed using nine 10 ms Hamming windows, located in each fricative as follows:

Ensemble Averaging



Time Averaging



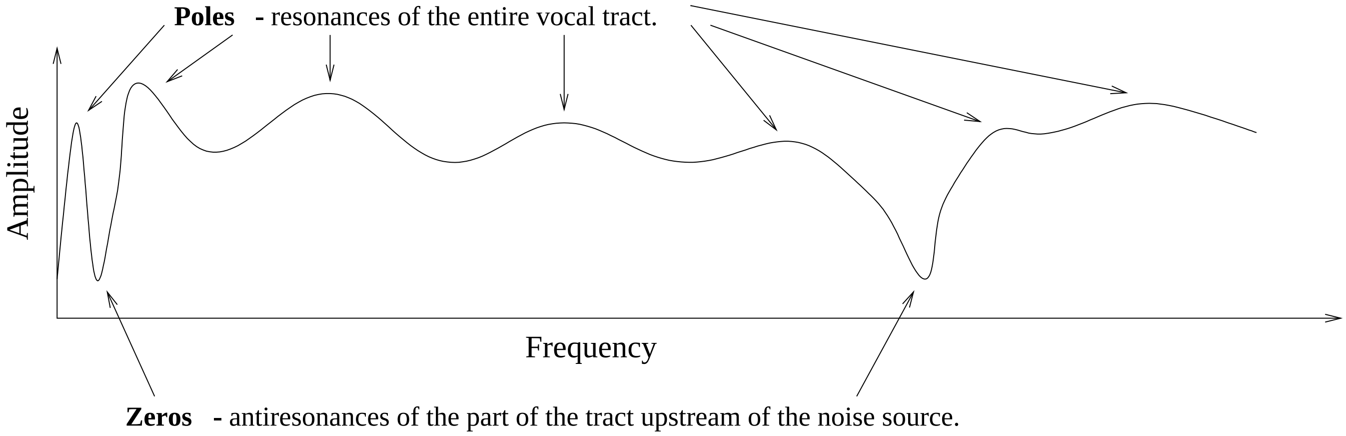
PARAMETERIZATION

The parameters are applied to the far-field spectrum, and consist of measures of the dynamic range of the spectrum, and spectral slope.

First defined from mechanical model results (Shadle, 1985), and further developed for classifying fricatives from speech (Shadle and Mair, 1996).

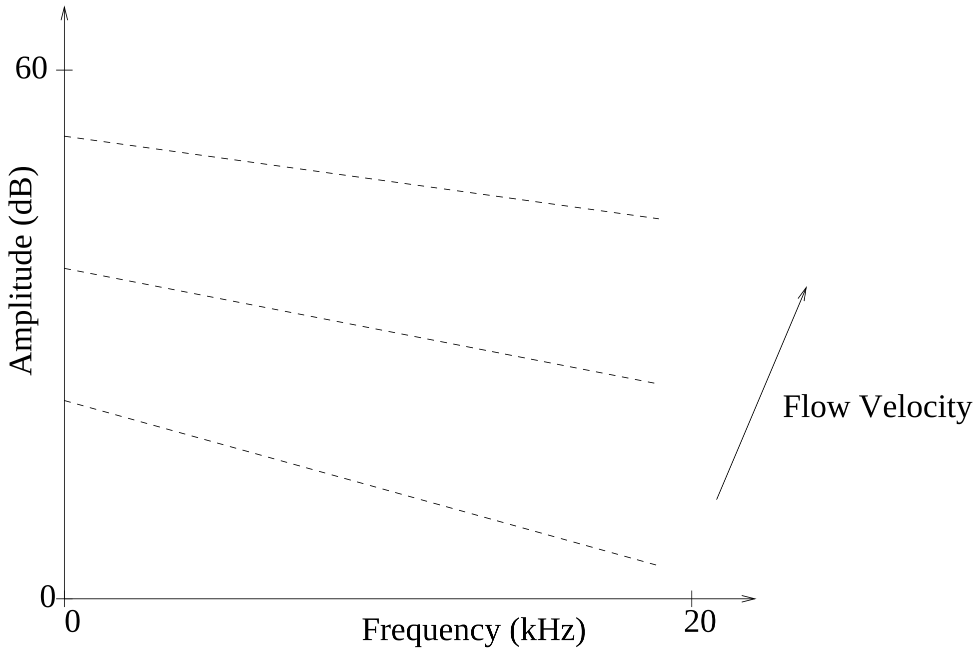
Far-field spectrum $\equiv$ excitation of the **tract transfer function** by: a source (for unvoiced) or sources (for voiced fricatives).

Transfer Function



The noise source spectrum depends on the shape of the constriction, the tract downstream of it, and the **flow velocity** through it.

Noise Source Spectral Envelope



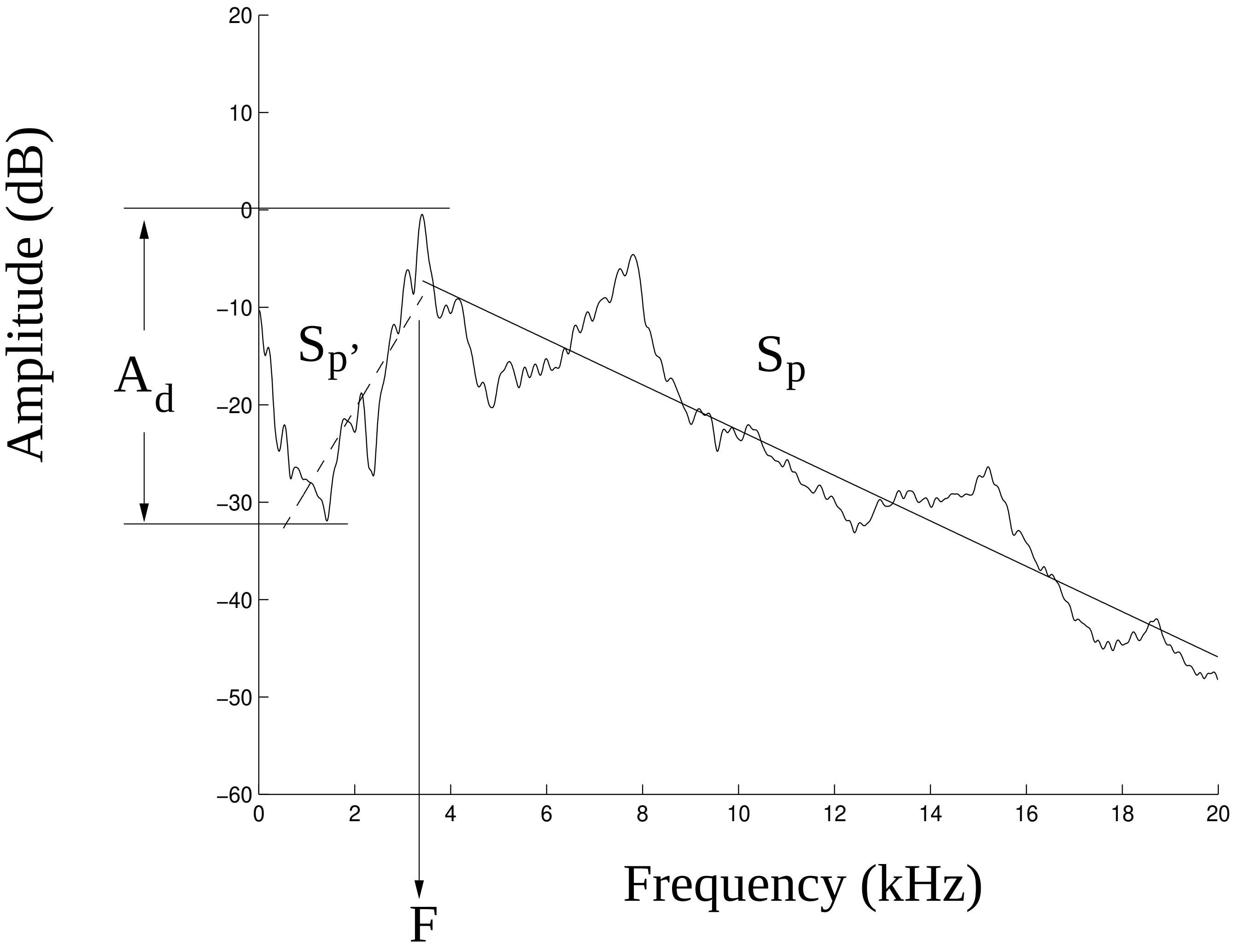
PARAMETERIZATION

Goal 1: identification of the fricative spoken regardless of its context or the way in which it was spoken ⇒
information mainly in the transfer function, and in the source type.

Goal 2: to describe the acoustic variation caused by the context or the way in which a particular fricative is spoken ⇒
information mainly in the source spectrum.

In this study we are primarily interested in **Goal 2**.

DEFINITION OF PARAMETERS



Sustained fricative /ʃ/ (**Corpus 1a**) produced by **Speaker ISSS**.

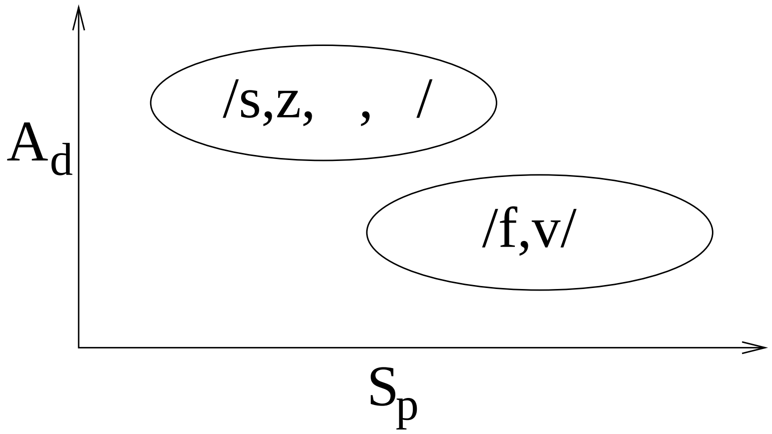
1. Determine F, the frequency of the spectral peak between 2 and 8 kHz having maximum amplitude, which corresponds to the same cavity resonance for all tokens of a particular fricative.
2. Find $A_d \equiv$ difference between the maximum amplitude between 0.5 and 20 kHz, and the minimum amplitude between 0 and 2 kHz.
3. Fit regression line from 500 to F Hz $\Rightarrow$ slope of line $\equiv S_{p'}$.
4. Fit regression line from F to 20 kHz $\Rightarrow$ slope of line $\equiv S_p$.

PREDICTED EFFECTS ON PARAMETERS

Phonetic Class	Aeroacoustics	Predictions
Posterior place; sibilants /s,z,ʃ,ʒ/	Longer front cavity; Localized source; Higher source strength *	F lower; A_d, S_p and $S_{p'}$ higher
Forward place; nonsibilants /f,v/	Distributed source; Lower source strength	F higher; A_d, S_p and $S_{p'}$ lower
Unvoiced	Higher source strength *	A_d, S_p and $S_{p'}$ higher
Voiced-Devoiced		
Voiced	Lower source strength	A_d, S_p and $S_{p'}$ lower
Loud effort level	Higher source strength *	A_d, S_p and $S_{p'}$ higher
Medium effort level		
Soft effort level		
Beginning of fricative		
Middle of fricative	Higher source strength *	A_d, S_p and $S_{p'}$ higher
End of fricative		
Stressed syllable	Higher source strength *	A_d, S_p and $S_{p'}$ higher
Destressed syllable		
Word position		
Rounded	Longer front cavity; Lower source strength	F lower; A_d higher; ? $S_{p'}$ and S_p lower
Unrounded		
Subject		? No effect

*A higher source strength means higher airflow for the same constriction area A_c, or a constant flow for a smaller A_c.

PREDICTIONS OF A_d VS. S_p



ʃ

ʃ

ʃ

ʃ

ʃ

ʃ

ʒ

ʒ

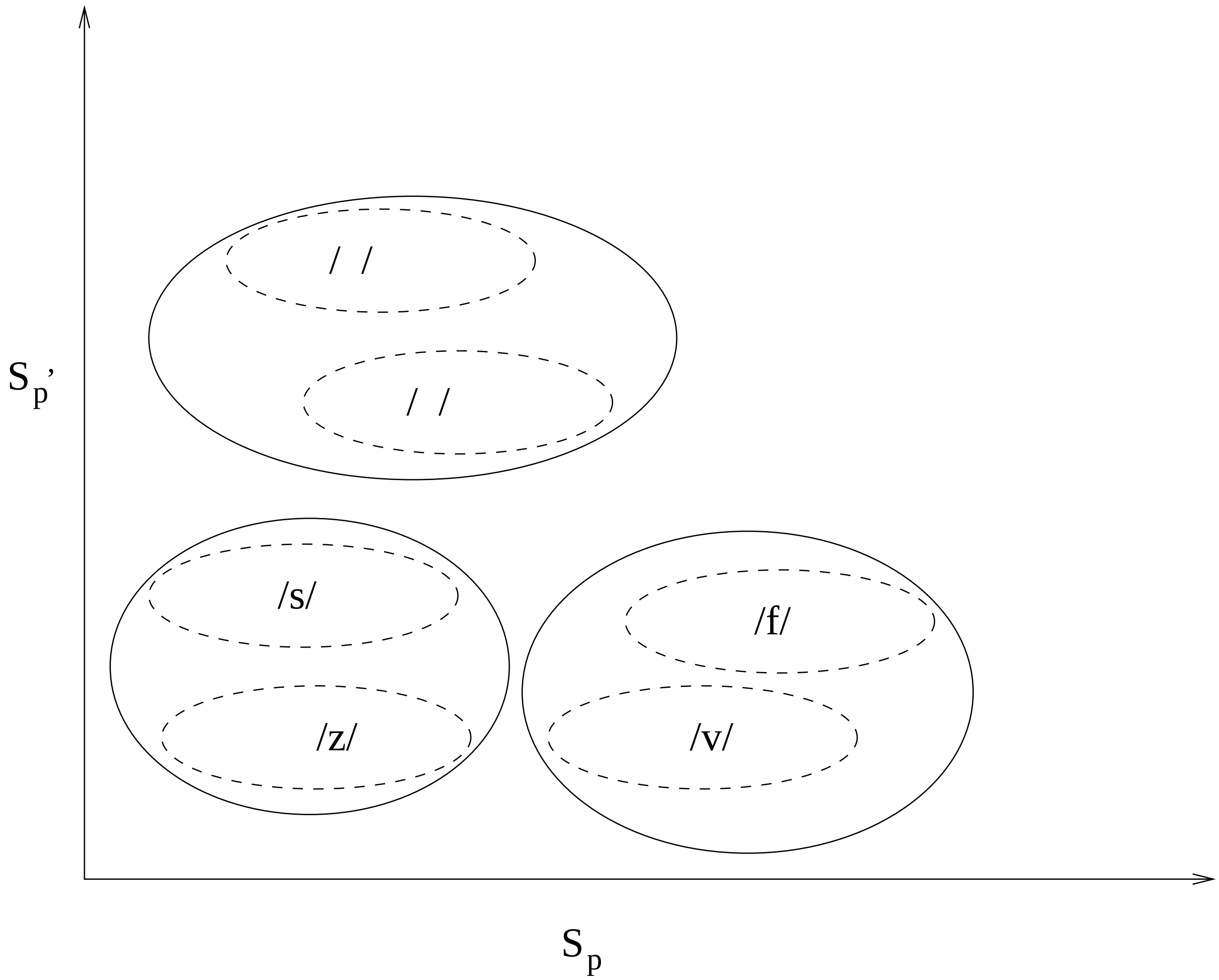
ʒ

ʒ

ʒ

ʒ

PREDICTIONS OF $S_{p'}$ VS. S_p



On an $S_{p'}$ vs. S_p plot, each place will cluster separately, with voiced tokens having lower $S_{p'}$ but similar S_p relative to their unvoiced counterparts.

RESULTS - Sustained Fricatives

- each place has a different "family" of nearly-parallel lines;
- higher effort level increases amplitude significantly and slope slightly;
- the families of lines for the voiced and unvoiced fricatives always overlap, with the voiced cases mostly lower in amplitude and occupying a smaller range of amplitudes than the unvoiced cases.

Average regression line fits (from F to 20 kHz) of sustained labiodental (top), alveolar (middle) and postalveolar (bottom) fricatives from **Corpus 1b** at loud, medium and soft effort levels. **Speaker ISSS**.

/f/

/v/

/s/

/z/

/ʃ/

/ʒ/

Amplitude (dB)

Frequency (kHz)

For A_d vs. fricative:

- all subjects, Corpus 1a and 1b, /s, z, ʃ, ʒ/ have a higher A_d than /f, v/;
- A_d differentiates between voiced fricatives and their unvoiced counterparts.

Dynamic amplitude of fricatives from **Corpus 1a**. **Speaker LMTJ**.

S_p vs. effort level, by fricative:

- slope generally increases with increased effort level, though this pattern is much more consistent for unvoiced fricatives. This is consistent with results in Shadle and Mair (1996).

Spectral slope of sustained fricatives from **Corpus 1b** at Loud (L), Medium (M) and Soft (S) effort levels. The horizontal line is the average value of all the examples. **Speaker ACC**.

RESULTS - Sustained Fricatives - A_d vs. S_p

- Sustained fricatives form two distinct clusters, of sibilants and /f,v/.

Corpus 1a (sustained fricatives), A_d vs. S_p.

RESULTS - Sustained Fricatives - $S_{p'}$ vs. S_p

- For speakers **LMTJ**, **CFGA** and **ISSS**, our results confirm the findings of Evers et al. (1998), i.e., that it is possible to separate /s/ from /ʃ/.
- For these subjects, fricatives /f,v/, /s,z/ and /ʃ,ʒ/ form clusters in the feature space, i.e., they are separated by place.
- The voiced tokens of each had lower $S_{p'}$ and similar S_p than their unvoiced correlates.
- For **ACC** the voicing relationship was maintained, but /s,z/ tokens fell inbetween the /ʃ/ and /ʒ/ tokens.

Corpus 1a (sustained fricatives), $S_{p'}$ vs. S_p.

RESULTS - Fricatives in Context

- For /f,v/ there is no consistent pattern; results in Shadle, Mair and Carter (1996) indicate that the vowel context may play more of a role.
- A_d separates sibilants from /f, v/ as for the sustained fricatives.
- A_d is higher mid-fricative than at beginning, end, for /s, z, ʃ, ʒ/
- S_p is lower mid-fricative for /s, z, ʃ, ʒ/, not as predicted.

Dynamic amplitude (top) and spectral slope (bottom) of fricatives from **Corpus 2**, at the Beginning (B), Middle (M) and End (E) of the fricative. **Speaker LMTJ**.

Preliminary comparisons of stressed and destressed fricatives indicate little or no change in A_d and S_p, not as predicted.

Note:

- Syllable stress is strongly correlated with the amount of devoicing.
- Portuguese fricatives devoice $> 50\%$.
- For fricatives in destressed syllables we predict: $A_d \searrow$ and $S_p \searrow$.
- For fricatives in stressed syllables we predict: $A_d \nearrow$ and $S_p \nearrow$.
- There may be some interaction!

Corpus 3 - /ʃ/

Corpus 2 - /z/

Dynamic amplitude (top) and spectral slope (bottom). The horizontal line is the average value of all the examples. **Speaker LMTJ**.

5 CONCLUSIONS

The parameters spectral slope, frequency of maximum amplitude, and dynamic amplitude, were developed to characterize fricative spectra, and applied to corpora recorded by four native Portuguese speakers.

The parameters behaved as predicted for changes in effort level, voicing, and location within the fricative.

The parameters capture source-related changes for the most part **as predicted**; for the sustained fricatives, they also separate fricatives by place. However, for the nonsense words of corpus 2, comparisons of stressed and destressed fricatives indicate little or no change in A_d and S_p, **not as predicted**. Since this pattern occurs also in real words of Corpus 3, this may be a characteristic of Portuguese.

A combination of parameters A_d and S_p was also useful for separating the fricatives by sibilance, and a combination of parameters S_p and $S_{p'}$ separated the fricatives both by place and sibilance.

ACKNOWLEDGEMENTS

This work was partially supported by Fundação para a Ciência e a Tecnologia, Portugal.

REFERENCES

V. Evers, H. Reetz, A. Lahiri (**1998**). "Crosslinguistic Acoustic Categorization of Sibilants Independent of Phonological Status", Journal of Phonetics, **26**, 345-370.

K. Forrest, G. Weismer, P. Milenkovic and R. Dougall (**1988**). "Statistical Analysis of Word-Initial Voiceless Obstruents: Preliminary Data", JASA, **84**, 115-123.

L. Jesus and C. Shadle (**1999**). "Acoustic Analysis of a Speech Corpus of European Portuguese Fricative Consonants", EuroSpeech'99, **1**, Budapest, Hungary, 431-434.

A. Jongman and J. Sereno (**1995**). "Acoustic Properties of Non-Sibilant Fricatives", ICPhS 95, **4**, Stockholm, Sweden, 432-435.

A. Jongman, J. Sereno, R. Wayland and S. Wong (**1998**). "Acoustic Properties of English Fricatives", 16th ICA, **4**, Seattle, USA, 2935-2936.

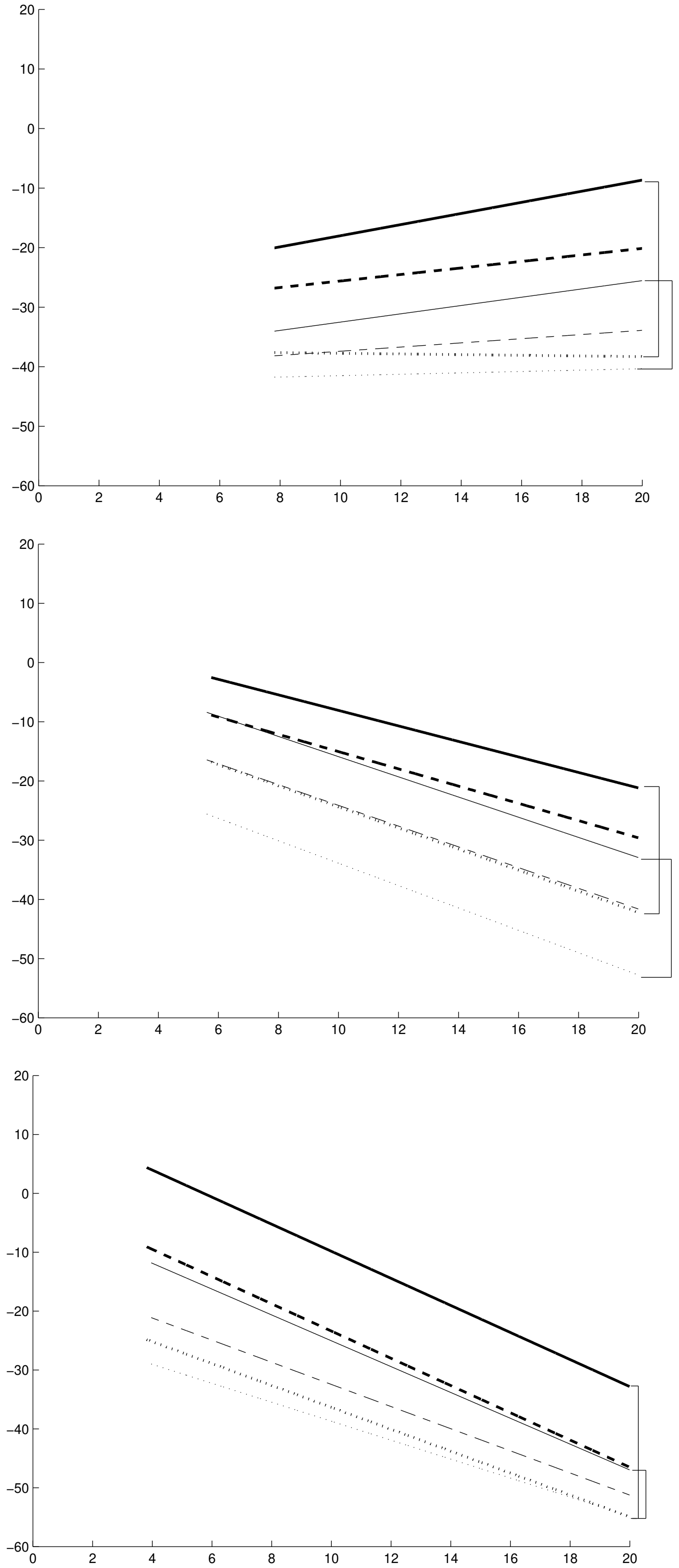
C. Shadle (**1985**). "The Acoustics of Fricative Consonants", Ph.D. Thesis, MIT, Cambridge, USA.

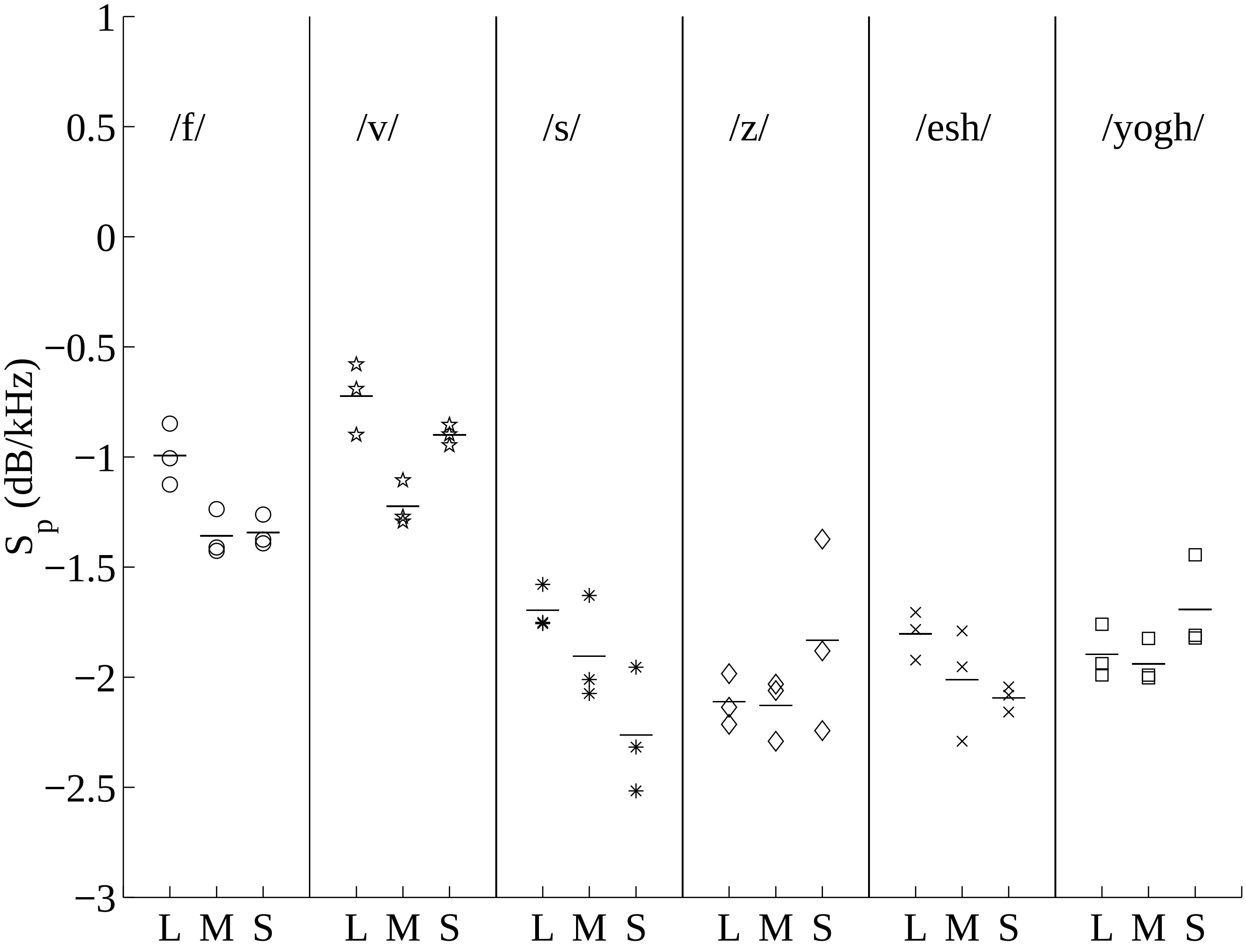
C. Shadle and S. Mair (**1996**). "Quantifying Spectral Characteristics of Fricatives". ICSLP 96, Philadelphia, USA, 1517-1520.

C. Shadle, S. Mair and J. Carter (**1996**). "Acoustic Characteristics of the Front Fricatives [f, v, θ, ð]". 4th Speech Production Seminar, Autrans, France, 193-196.

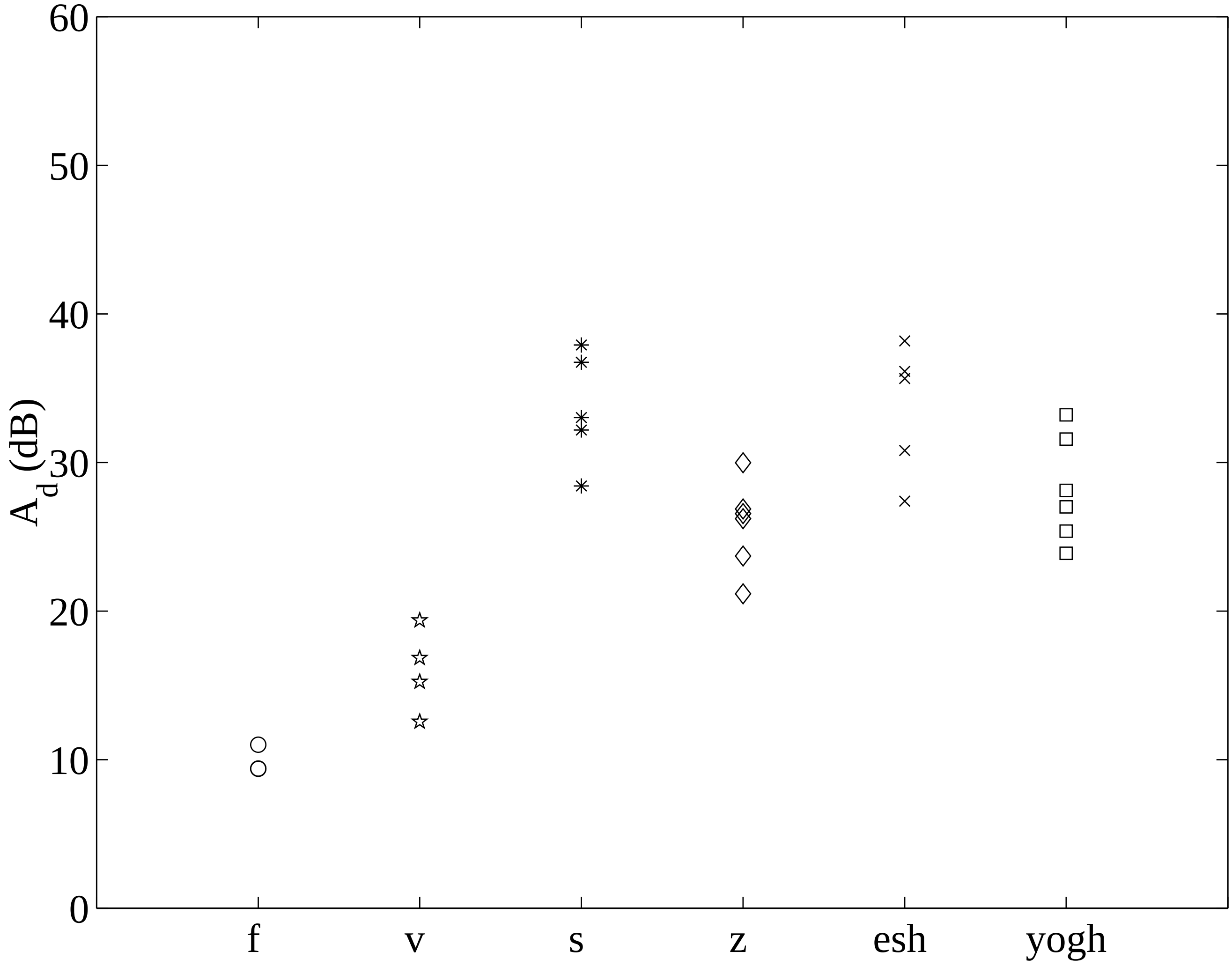
H. Sussman (**1994**). "The Phonological Reality of Locus Equations Across Manner Class Distinctions: Preliminary Observations", Phonetica, **51**, 119-131.

L. Wilde (**1995**). "Analysis and Synthesis of Fricative Consonants", Ph.D. Thesis, MIT, Cambridge, USA.

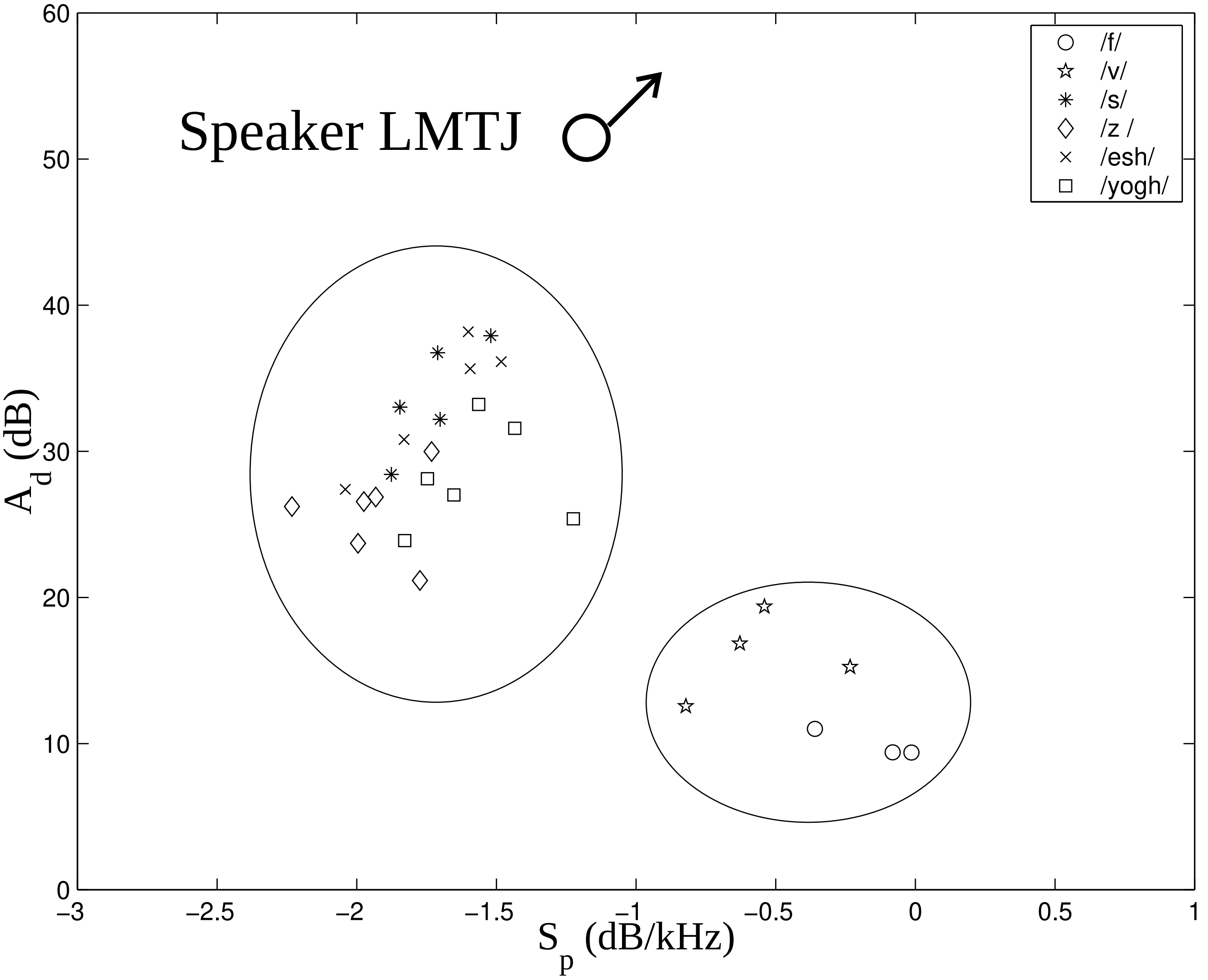

20
10
0
−10
−20
−30
−40
−50
−60
0 2 4 6 8 10 12 14 16 18 20
20
10
0
−10
−20
−30
−40
−50
−60
0 2 4 6 8 10 12 14 16 18 20
20
10
0
−10
−20
−30
−40
−50
−60
0 2 4 6 8 10 12 14 16 18 20



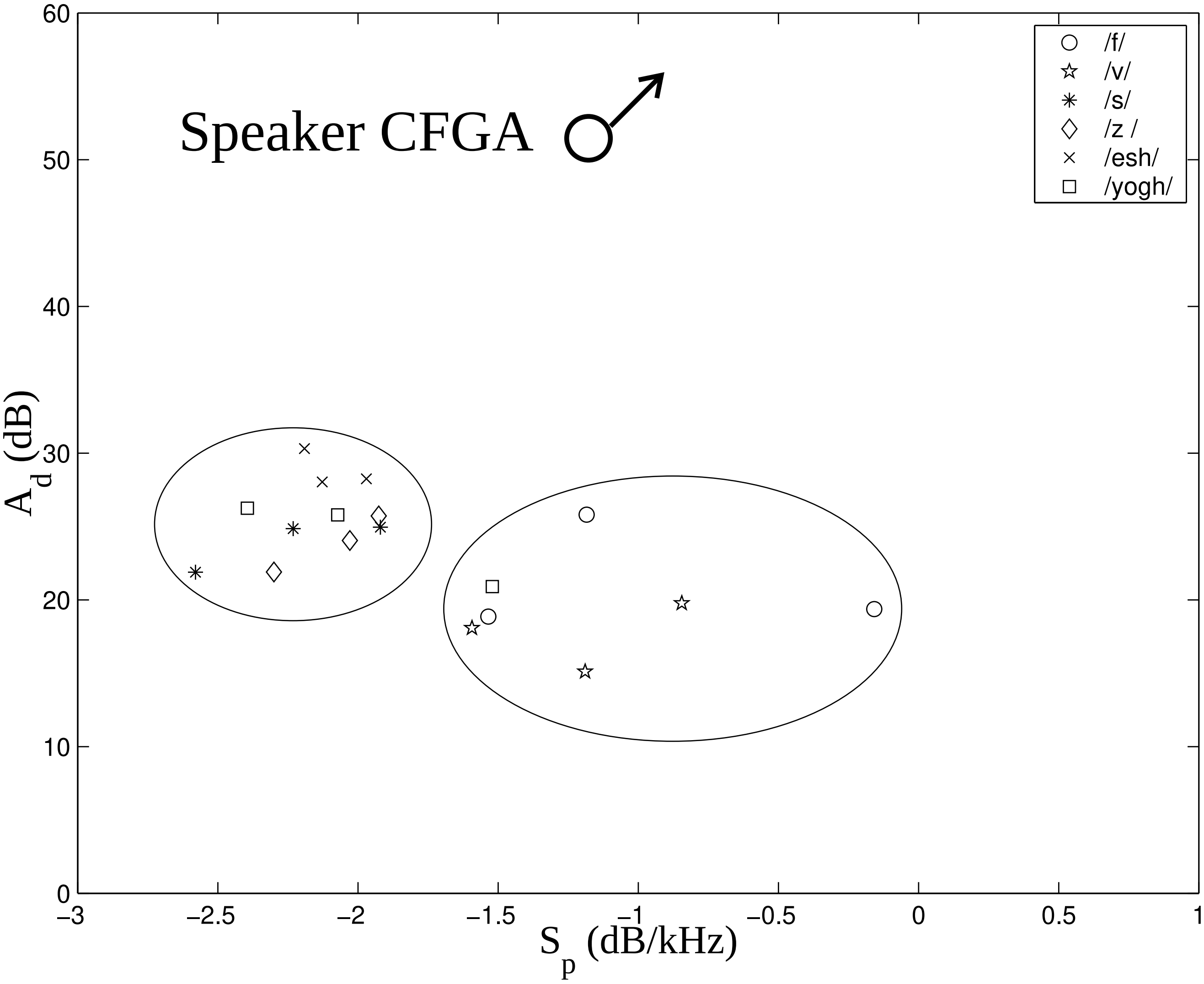
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−3
S_p (dB/kHz)
/f/
/v/
/s/
/z/
/esh/
/yogh/
L M S
L M S
L M S
L M S
L M S
L M S



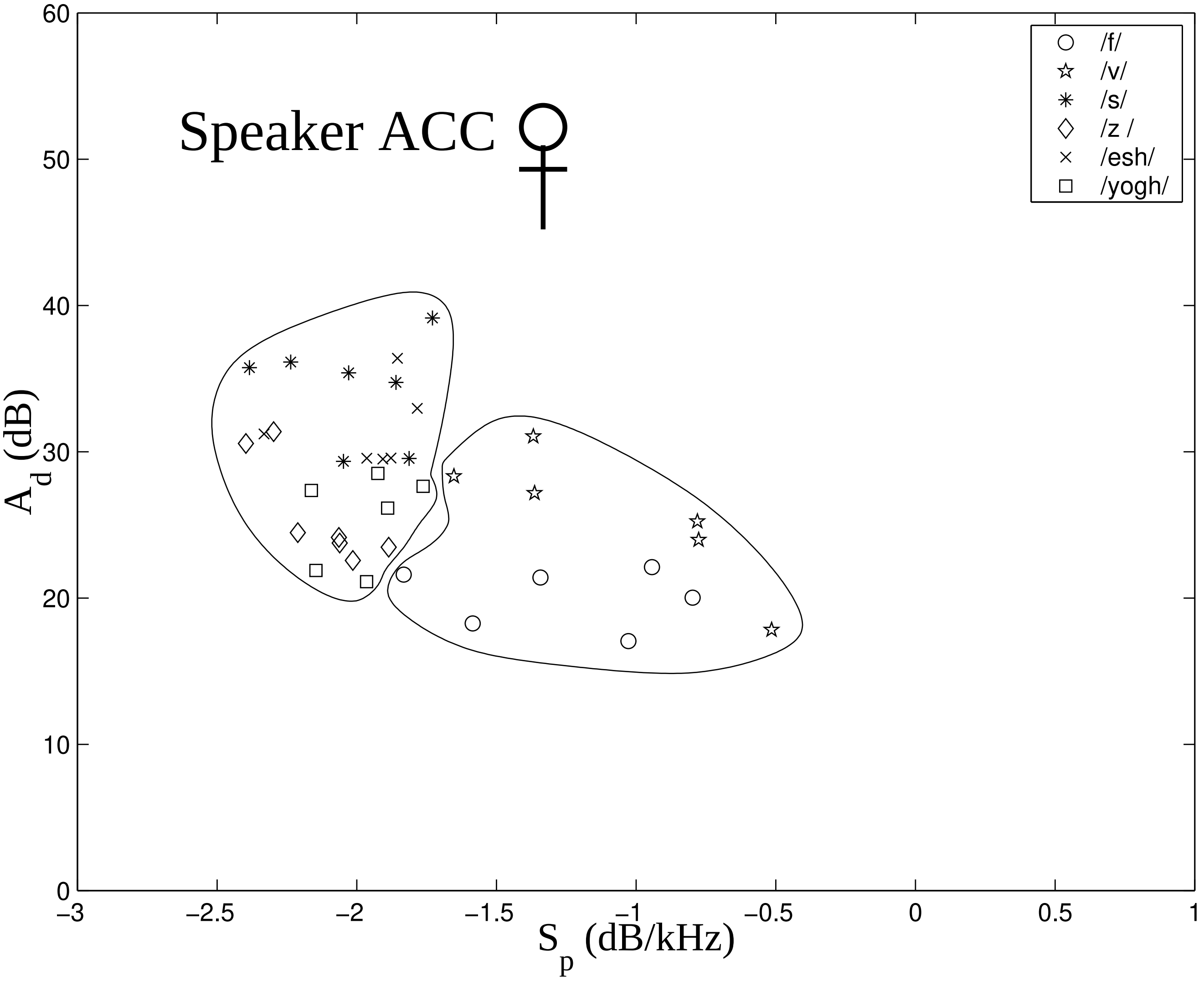
60
50
40
30
20
10
0
A_d (dB)
f
v
s
z
esh
yogh

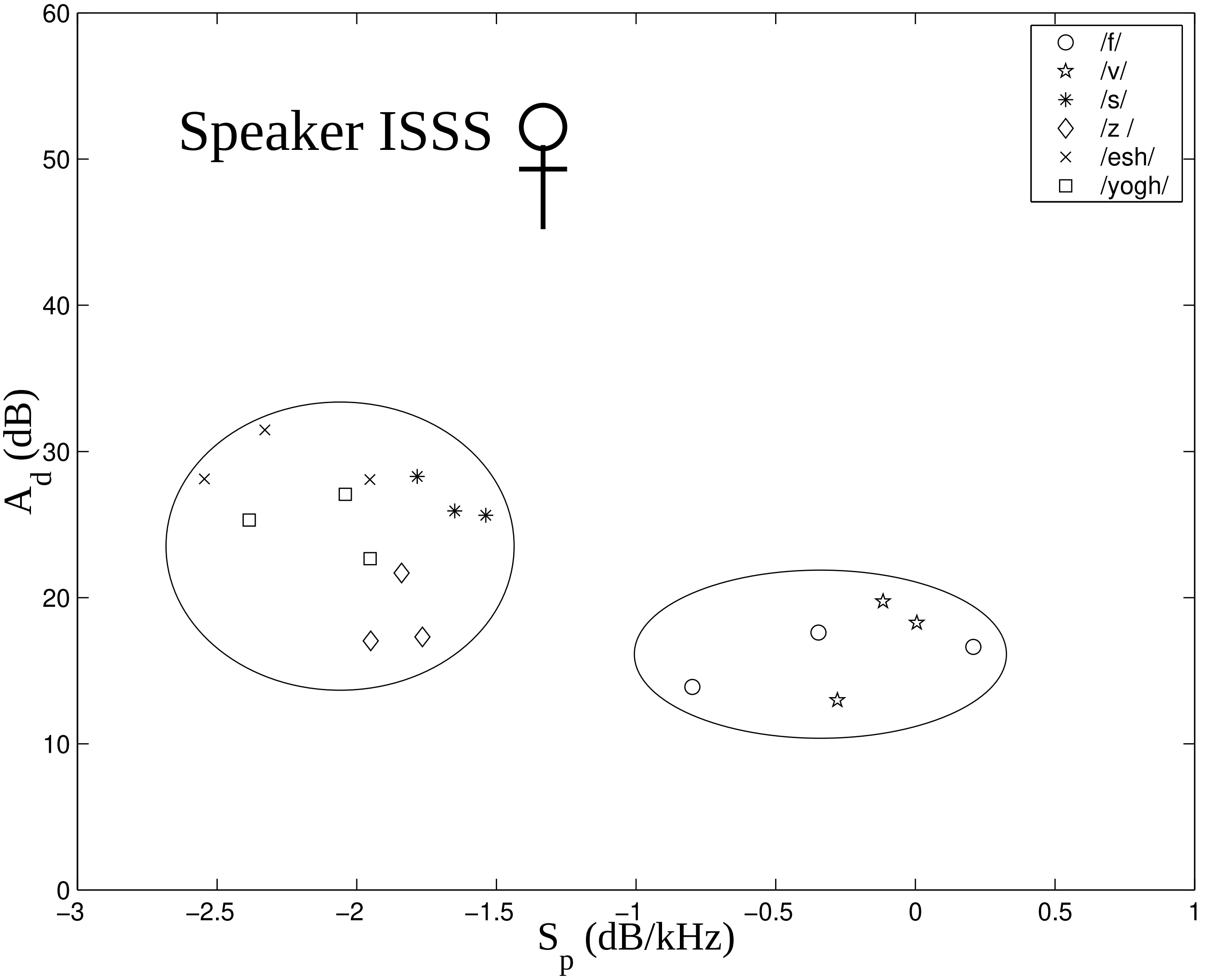


Speaker LMTJ
/f/
/v/
/s/
/z /
/esh/
/yogh/
60
50
40
30
20
10
0
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
A d (dB)
S p (dB/kHz)

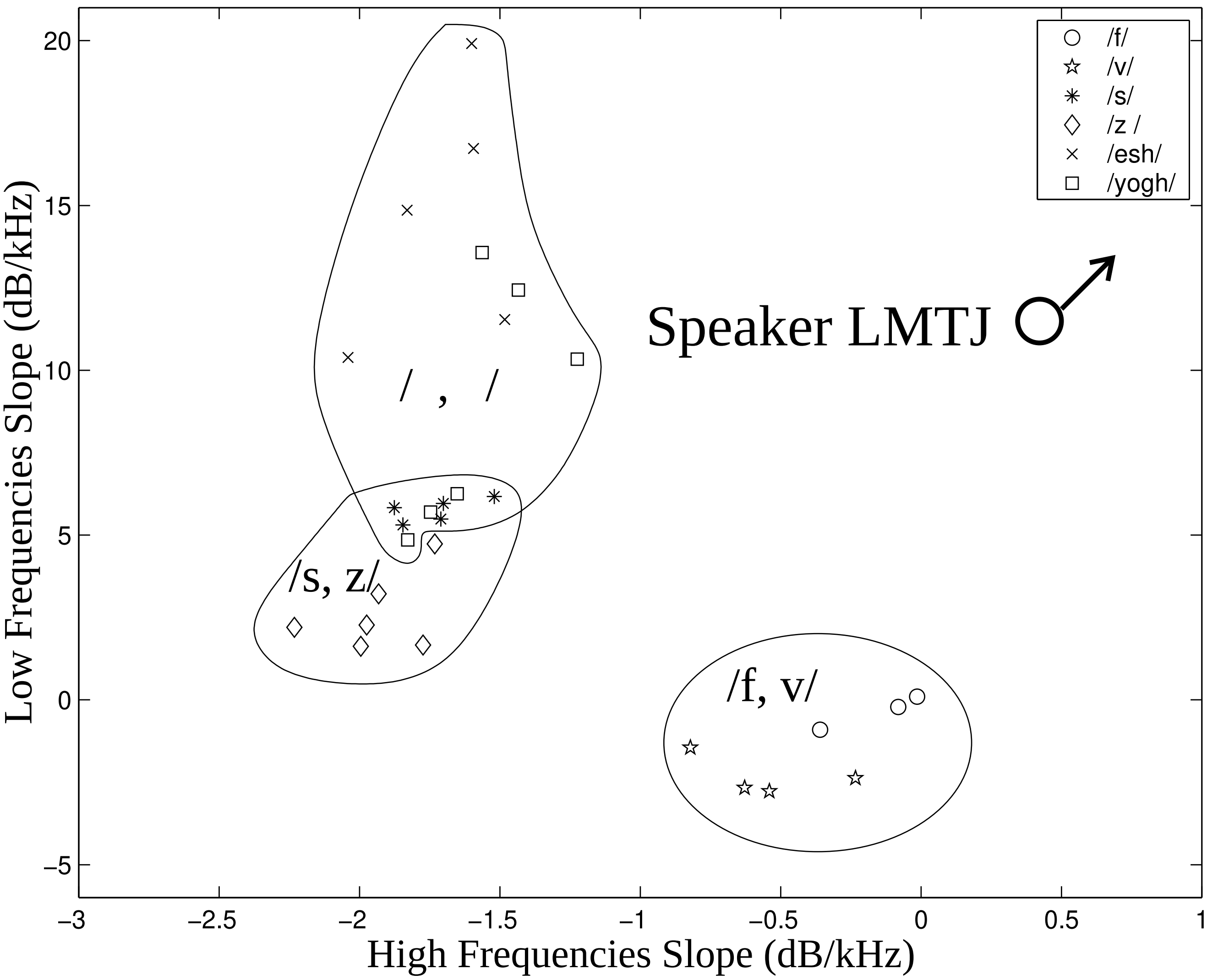


Speaker CFGA ♂
/f/
/v/
/s/
/z /
/esh/
/yogh/
A_d (dB)
S_p (dB/kHz)
0
10
20
30
40
50
60
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1

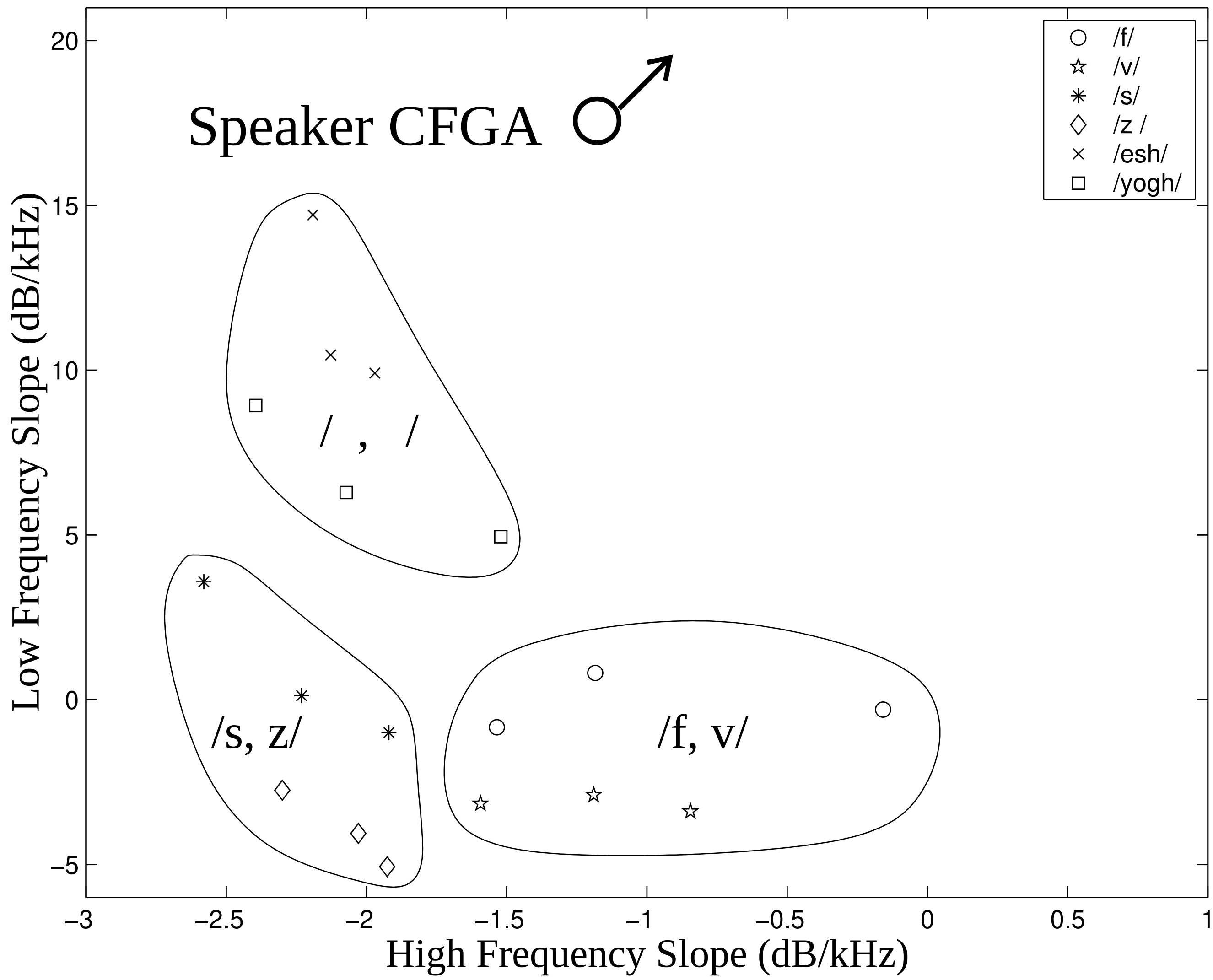

Speaker ACC ♀
/f/
/v/
/s/
/z /
/esh/
/yogh/
A_d (dB)
S_p (dB/kHz)
60
50
40
30
20
10
0
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1



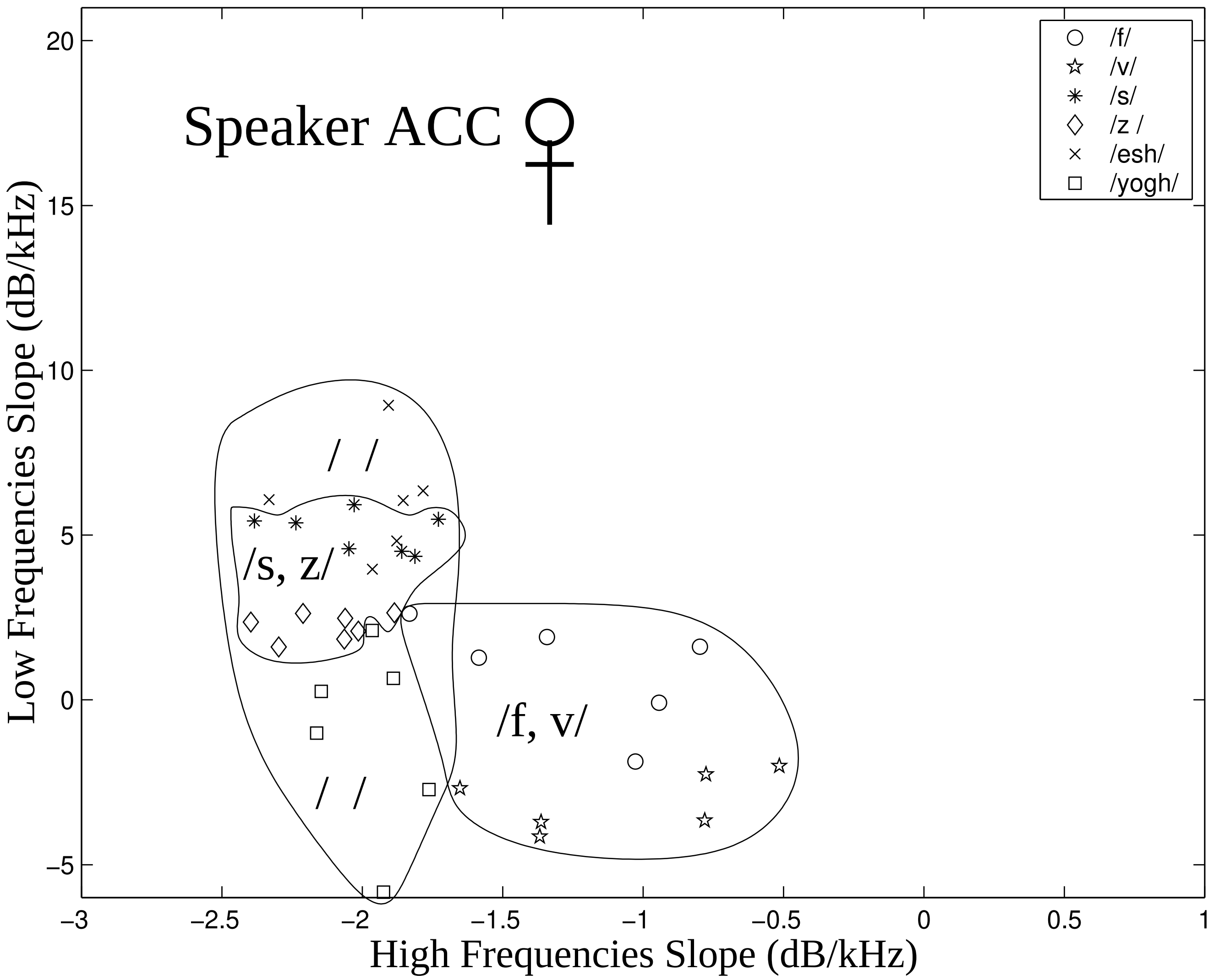
Speaker ISSS ♀
/f/
/v/
/s/
/z /
/esh/
/yogh/
A_d (dB)
S_p (dB/kHz)
0
10
20
30
40
50
60
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1



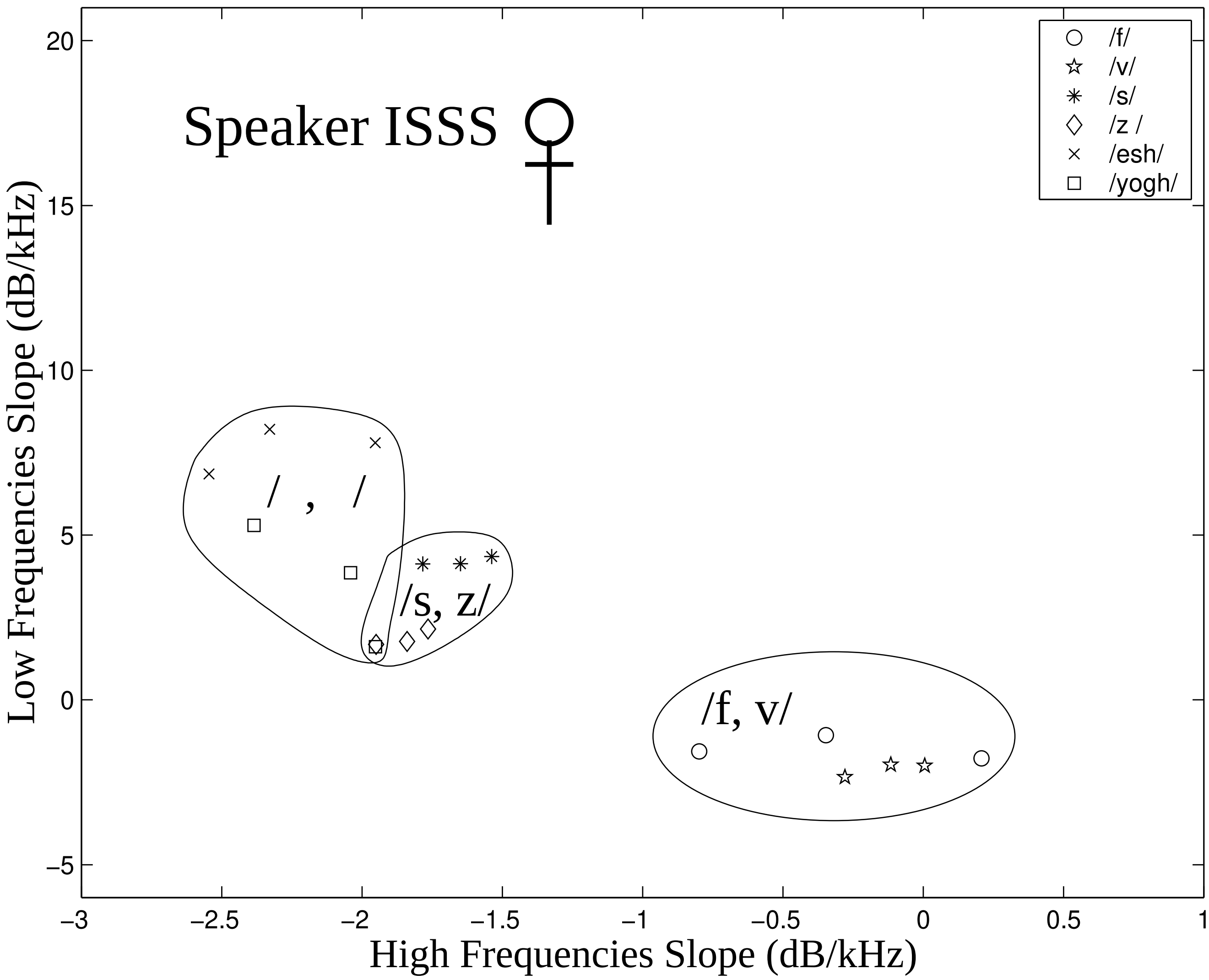
Speaker LMTJ ♂
/f/
/v/
/s/
/z /
/esh/
/yogh/
/ , /
/s, z/
/f, v/
Low Frequencies Slope (dB/kHz)
High Frequencies Slope (dB/kHz)
20
15
10
5
0
−5
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1

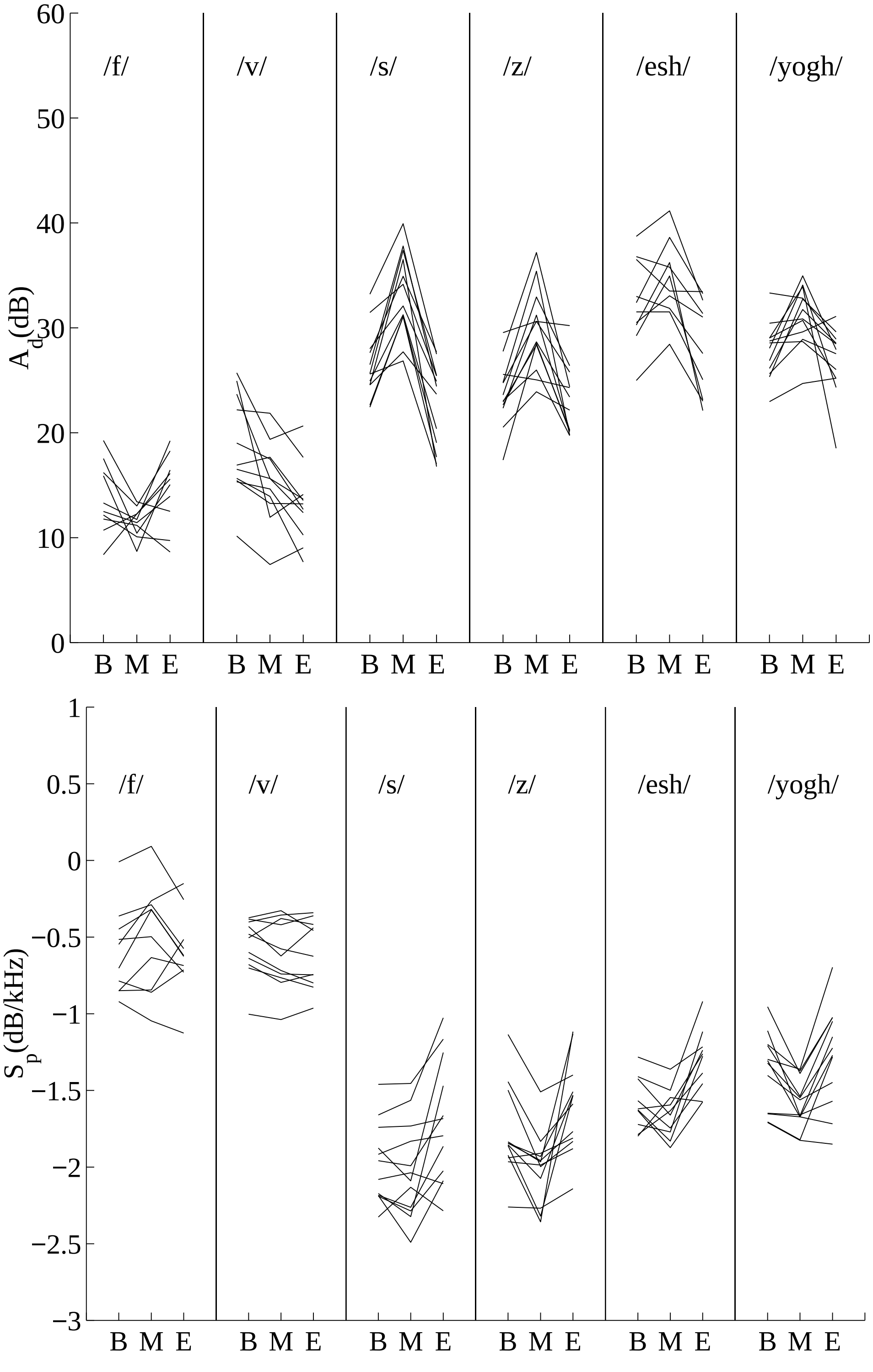


Speaker CFGA
/f/
/v/
/s/
/z /
/esh/
/yogh/
/ , /
/s, z/
/f, v/
20
15
10
5
0
−5
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
High Frequency Slope (dB/kHz)
Low Frequency Slope (dB/kHz)

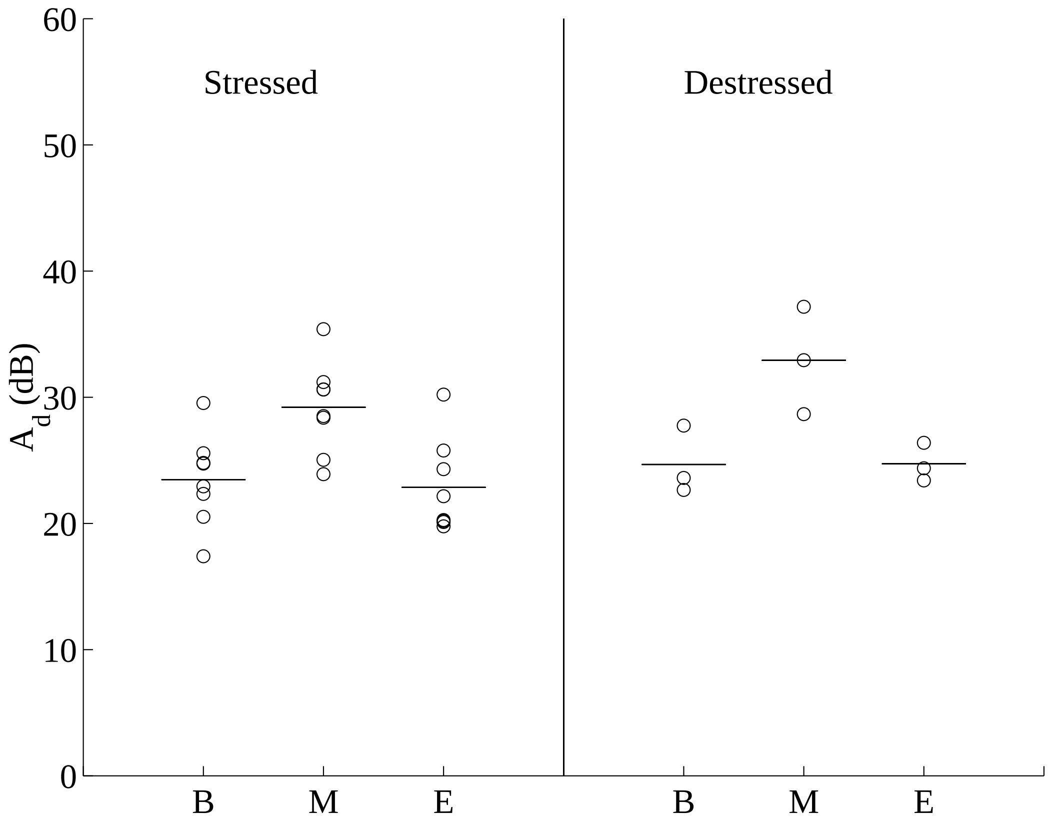


Speaker ACC ♀
/f/
/v/
/s/
/z /
/esh/
/yogh/
/ /
/s, z/
/f, v/
/ /
Low Frequencies Slope (dB/kHz)
High Frequencies Slope (dB/kHz)
20
15
10
5
0
−5
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1

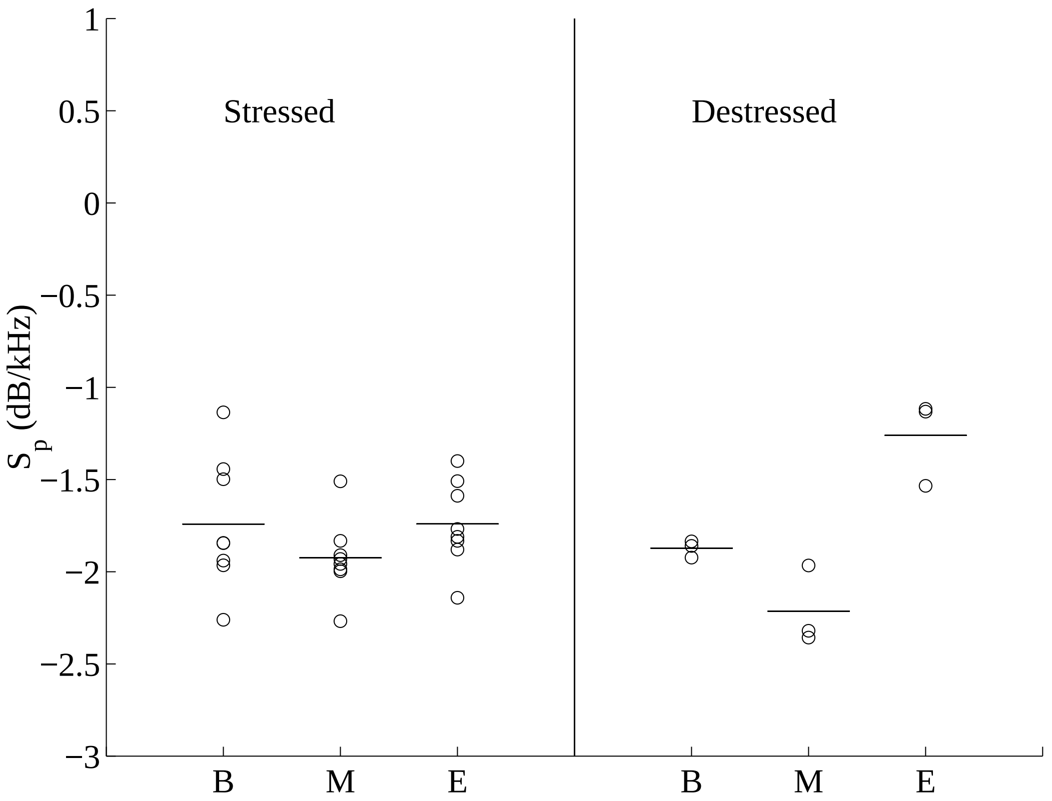

Speaker ISSS
/f/
/v/
/s/
/z /
/esh/
/yogh/
/ , /
/s, z/
/f, v/
Low Frequencies Slope (dB/kHz)
High Frequencies Slope (dB/kHz)
20
15
10
5
0
−5
−3
−2.5
−2
−1.5
−1
−0.5
0
0.5
1



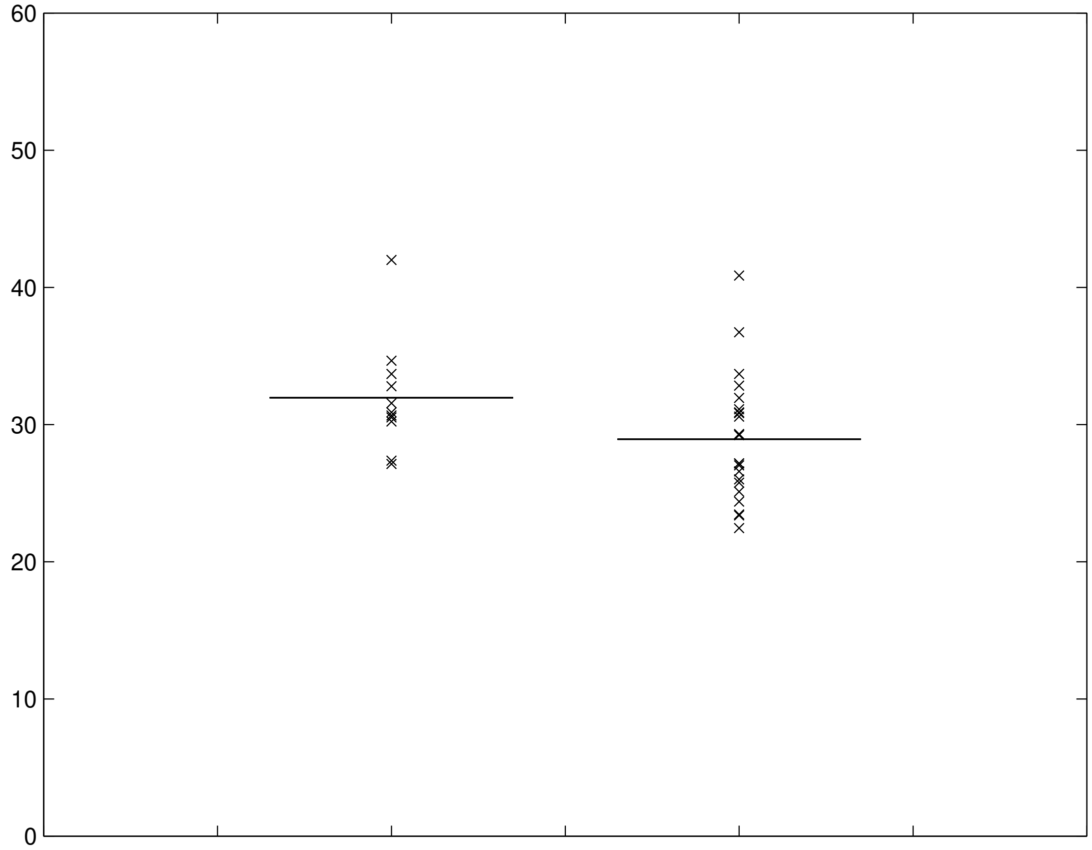
/f/
/v/
/s/
/z/
/esh/
/yogh/
A_d(dB)
0
10
20
30
40
50
60
B M E
S_p(dB/kHz)
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−3

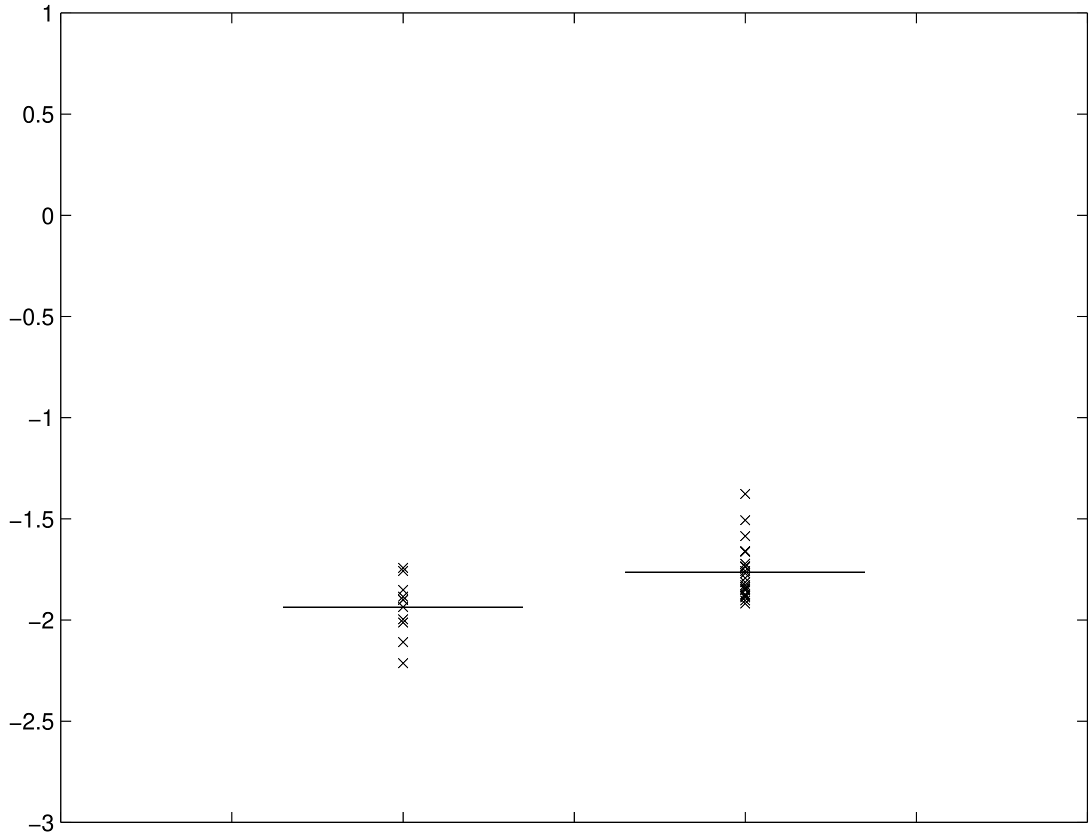


60
50
40
30
20
10
0
A_d (dB)
Stressed
Destressed
B
M
E
B
M
E



1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−3
S_p (dB/kHz)
Stressed
Destressed
B
M
E
B
M
E


60
50
40
30
20
10
0


1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−3