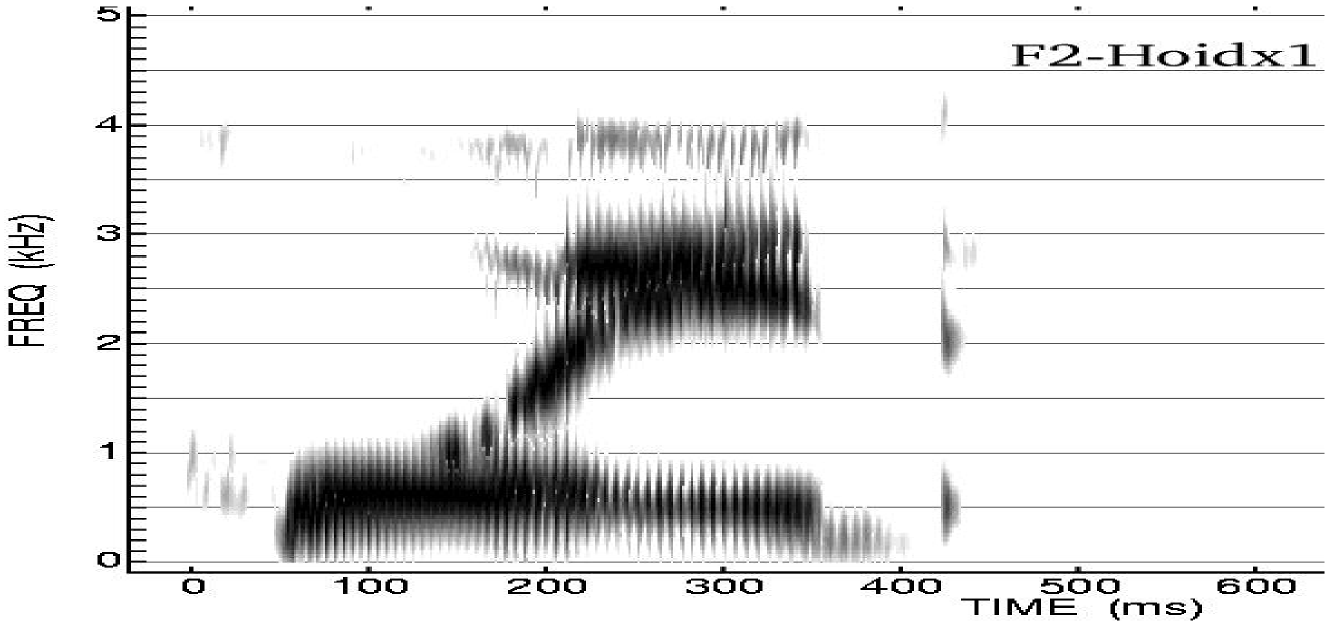
Subglottal Coupling and Vowel Space

Morgan Sonderegger &
Xuemin Chi
[smore, xuemin]@mit.edu
Speech Communication Group, MIT

Massachusetts
Institute of
Technology

rLe
AT MIT

Introduction

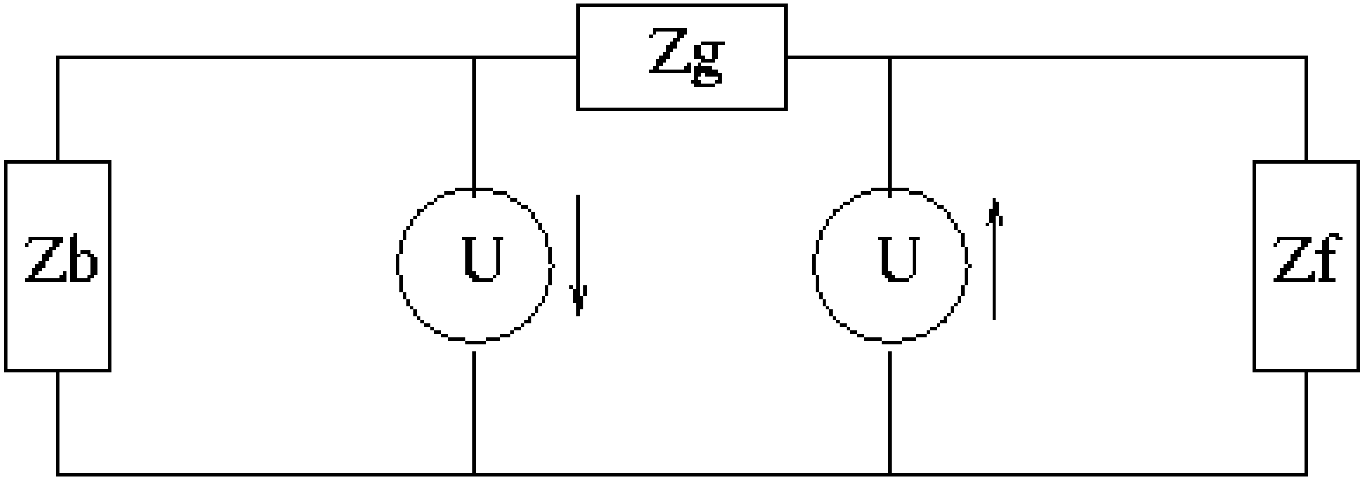


- In front-back diphthongs, observe attenuation of F2 peak at the 2nd subglottal resonance (AccF2).

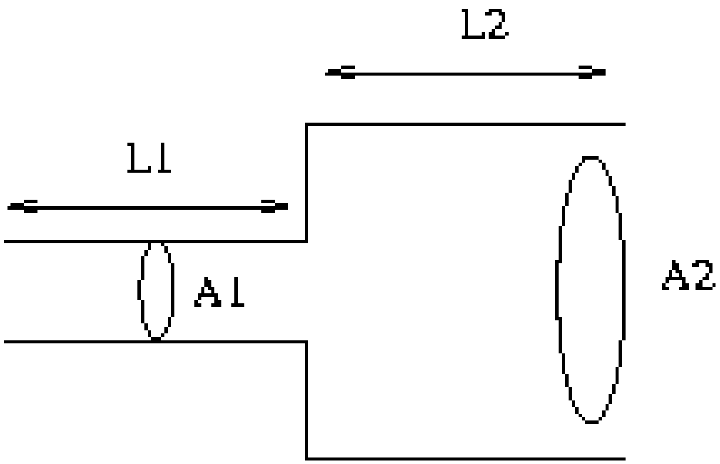
- AccF2 ~1350 Hz for males, ~1550 Hz for females.
- Front vowel F2 generally above and back vowel F2 generally below this frequency.
- What divides front and back vowels is uncertain.
- Hypothesis: attenuation is a quantal (Stevens 1989) phenomenon for [back].
- We model this effect, then test if it is quantal in several ways.

Theory

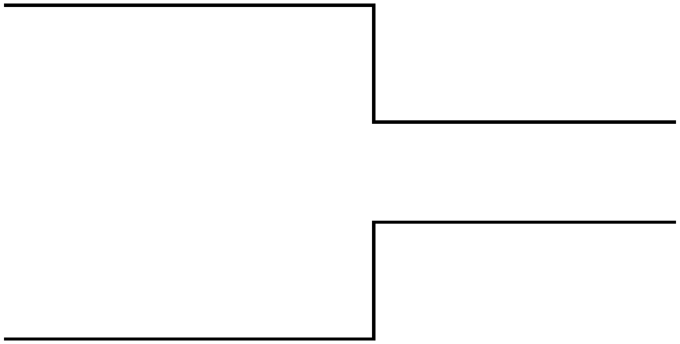
- Oral and subglottal cavities coupled at the glottis, impedances Z_f, Z_b, Z_g .
- What happens as F2 goes through a resonant frequency of the subglottal system?



- Subglottal system modeled as open tube terminated in lossy compliance.
- Oral cavity modeled as two tubes, sufficiently accurate for vowels.



/a/: A1/A2=1/8,
L1/L2=1.2



/i/: A1/A2=8,
L1/L2=1.5

- Wall impedances not included.
- $\text{Pressure at microphone} = \frac{\partial U}{\partial t}(\omega) \cdot T(\omega) \cdot R(\omega)$
- Get normal supraglottal poles, subglottal pole-zero pair.
- Pole-zero pair separation depends on oral-subglottal coupling (Z_g).
- Using model, can simulate attenuation in /ai/ diphthong (movie on author’s laptop):

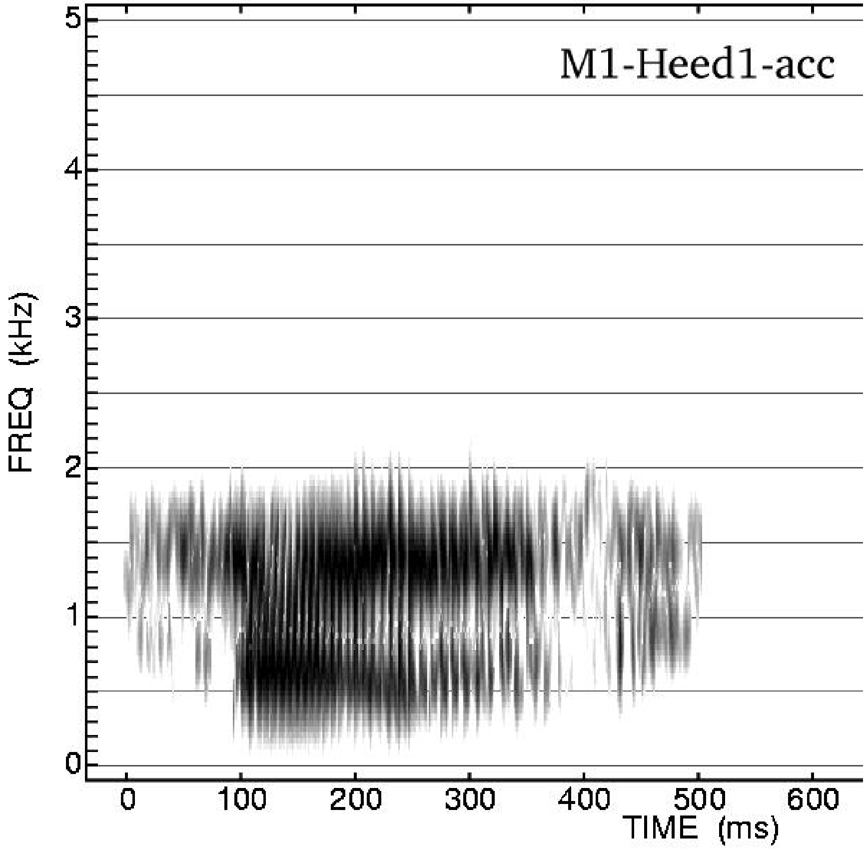
Data Collection

- Acoustic, accelerometer data to test hypothesis for individual speakers.
- 7 female, 6 male speakers.
- Native speakers of American English
- “hVd, say hVd again “, 5x, for all vowels.
- Same done for British English, Polish, one male speaker each.

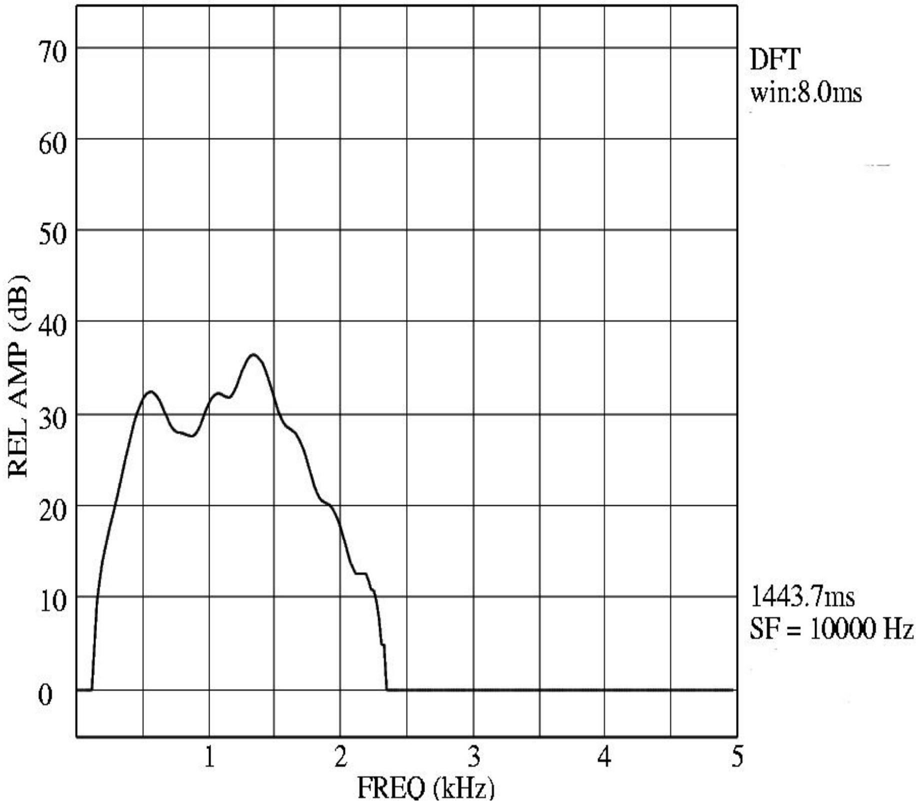
Accelerometer Details

- Glued to neck approximately 1 in. above the sternal notch. (Stevens et. al. 1975)
- Well-tested (Cheyne 1993), non-invasive.
- Converts the vibration of the skin to voltage signal => find subglottal resonance.

Sample Acc. Spectrogram



Spectrum of /heed/

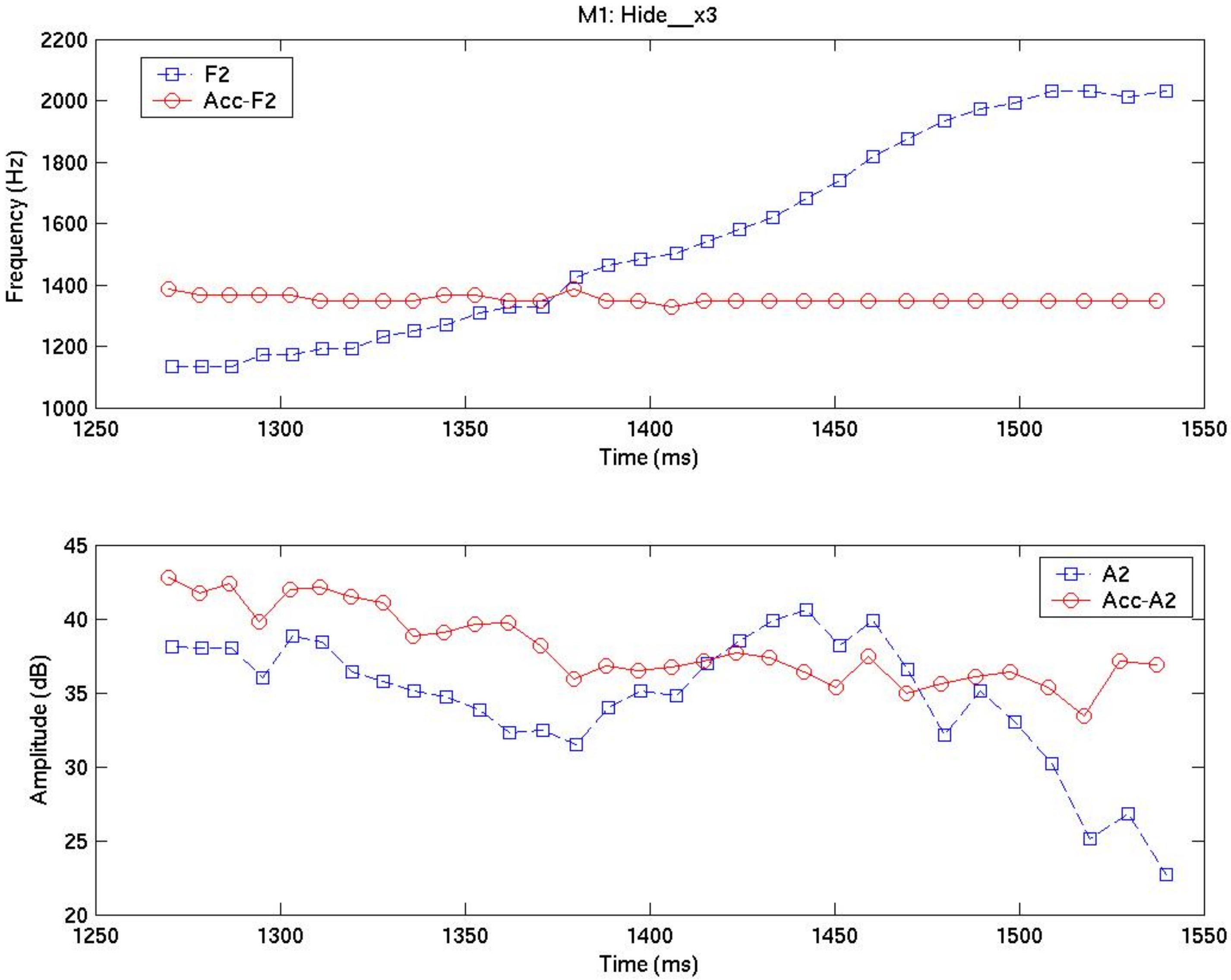


Speaker M1: acc-F1 = 547 Hz, AccF2 = 1360 Hz

Data Analysis-Diphthongs

- Looked at “hoid” and “hide”, in isolation for 3 male and 3 female speakers.
- Formants and amplitudes recorded by pitch period.
- See “jumps” in frequency for some speakers, not others, ~100-200 Hz.
- When there is a jump, amplitude dips, qualitatively matches the acoustic model.
- When there is no obvious jump, amplitude dips around AccF2, suggesting possibility of coupling.

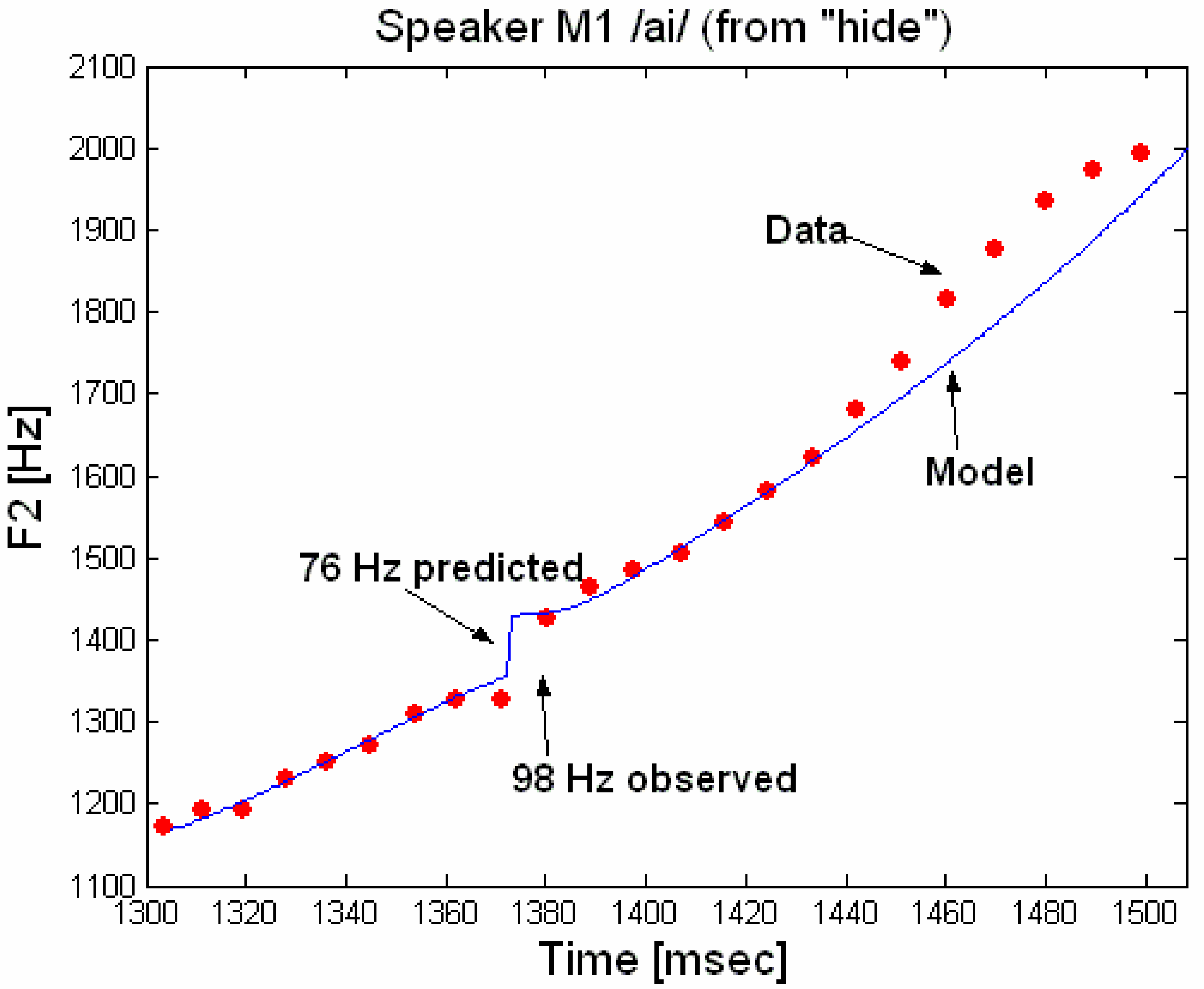
Data Showing a frequency jump



Data Analysis-Monophthongs

- Examined front vowels, F2 clearly above the measured 2nd sub-glottal resonance for all speakers.
- For back vowels, “hud” F2 is most often near Acc-F2, recorded “hub” to see if /d/ is pulling it up.
- “hub” F2<AccF2 for all but one speaker.
- Possible errors: accelerometer non-invasive, oral-subglottal coupling may shift measured resonance.

Modeling Jumps



- Can successfully model jumps.
- Model parameters can be adjusted to match magnitude, location of jump and attenuation.
- Non-robustness of effect predicted: too much or too little coupling gives no jump.
- Speakers’ jump characteristics vary with Z_b, $Z_{g.}$
- Shows that suggested quantal phenomena may not occur in practice, can predict via modeling.

Monophthong Statistics

- Are all front/back vowels above/below AccF2 for individual speakers?
- Subglottal resonance varies between utterances.
- ~160 vowels per speaker, found F2 by hand for all speakers.
- For each vowel, found mean AccF2 using a formant tracker.

Diversion: Acc-F2 Statistics

- AccF2 distribution for each individual speaker across all vowels is gaussian, $\chi_\nu^2 \sim 1$.
- Variance ~30-60 Hz
- No significant differences in AccF2 for different vowels=> AccF2 relatively stable for each speaker.
- Mean values:1280-1450 Hz for males, 1380-1620 for females.
- Agrees with work measuring AccF2 invasively (Cranen & Boves 1987, Ishizaka et. al. 1976).

Significance Testing

- Used all monopthongs for American, British speakers' dialects, plus /e/ from /ei/.
- For each speaker: for a given vowel, F2 error=variance of 10 F2 values (5 repetitions), the "vowel group."
- AccF2 error=variance of speaker's distribution.
- Tested whether AccF2-F2 significantly ($p<.05$) positive or negative for each group.

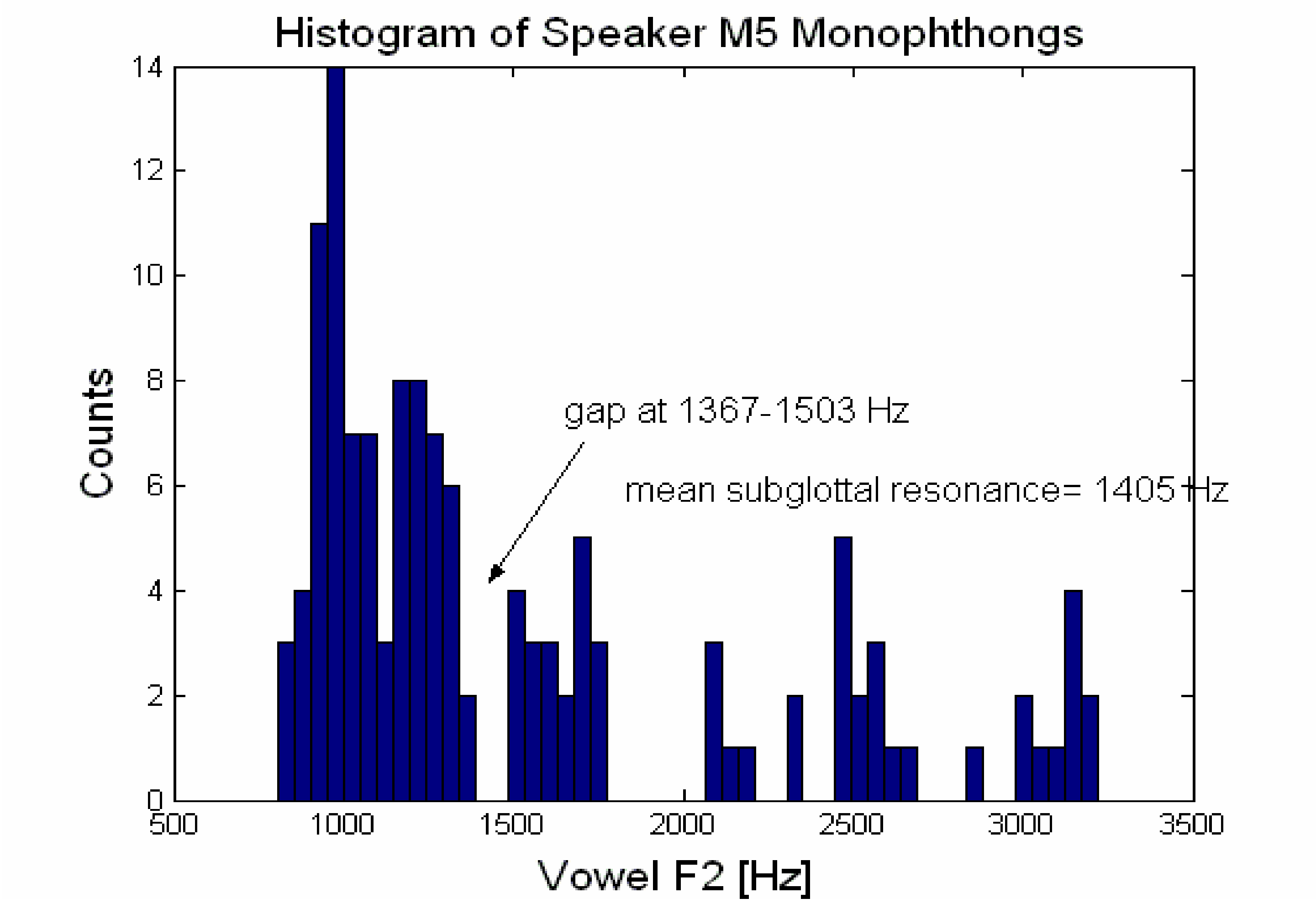
- Back vowel F2<Acc-F2 is “expected,” etc.
- 4 categories: significant & expected, non-significant & not expected, etc.
- Only groups for certain vowels ever not significant & expected: “hodd,” “hoed,” hood,” “hawed,” “hud,” “who’d.”
- Statistics for **these** vowels across 14 speakers, 78 groups:

Significant & expected	Non-significant and expected	Non-significant and non-expected	Significant and non-expected
65 (**86**%)	5 (6%)	1 (1%)	5 (6%)

- Front groups all > AccF2, few back groups problematic.
- Central (/er/) group above, below, or across AccF2 for different speakers.
- Pattern holds even for speakers without jumps.
- Using “hub” instead of “hud”:

Significant & expected	Non-significant and expected	Non-significant and non-expected	Significant and non-expected
65 (**91**%)	2 (3%)	1 (1%)	4 (5%)

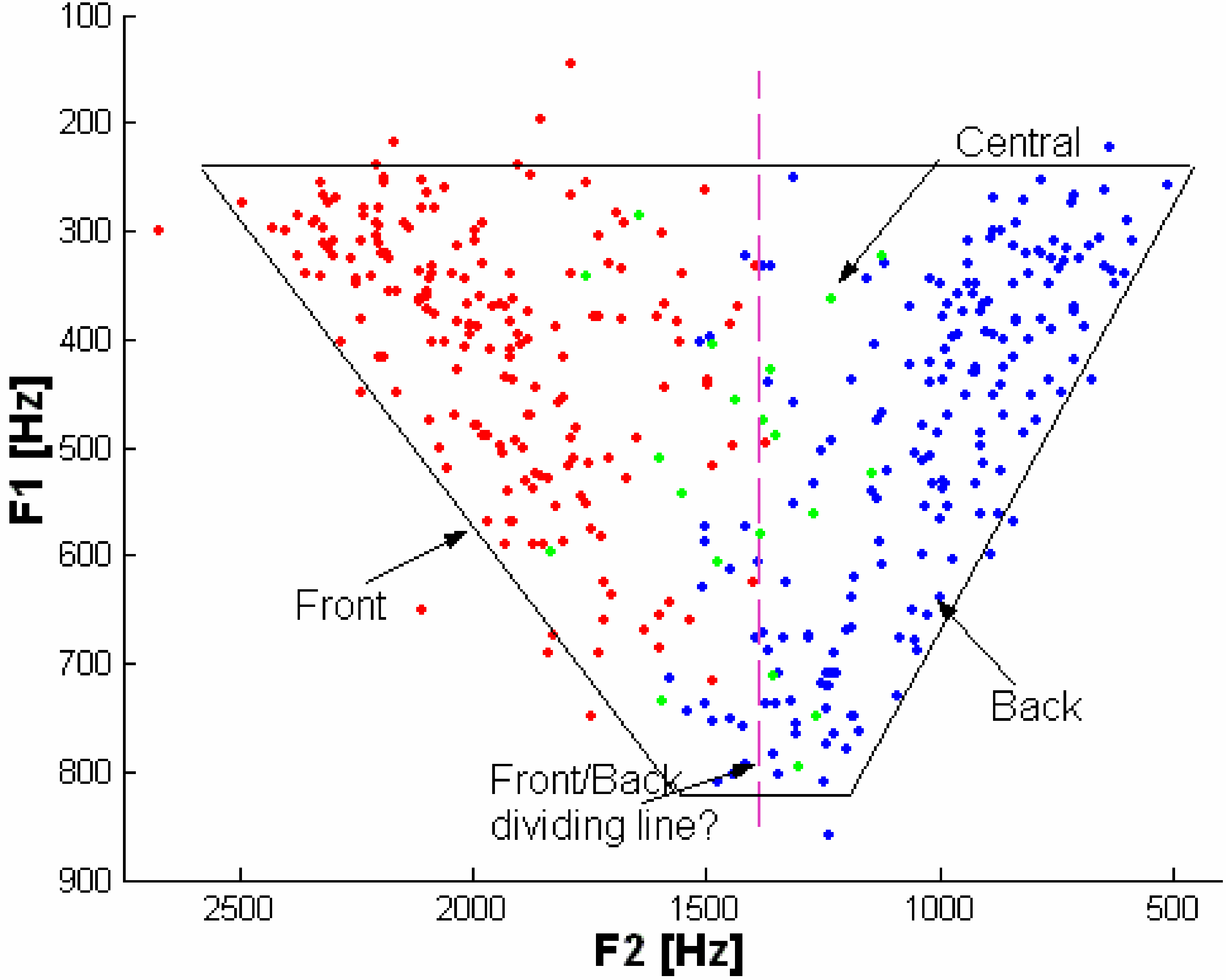
- Speakers with jumps have gaps in their F2 data for all vowels=> possible vowel spaces are constrained by attenuation phenomenon.



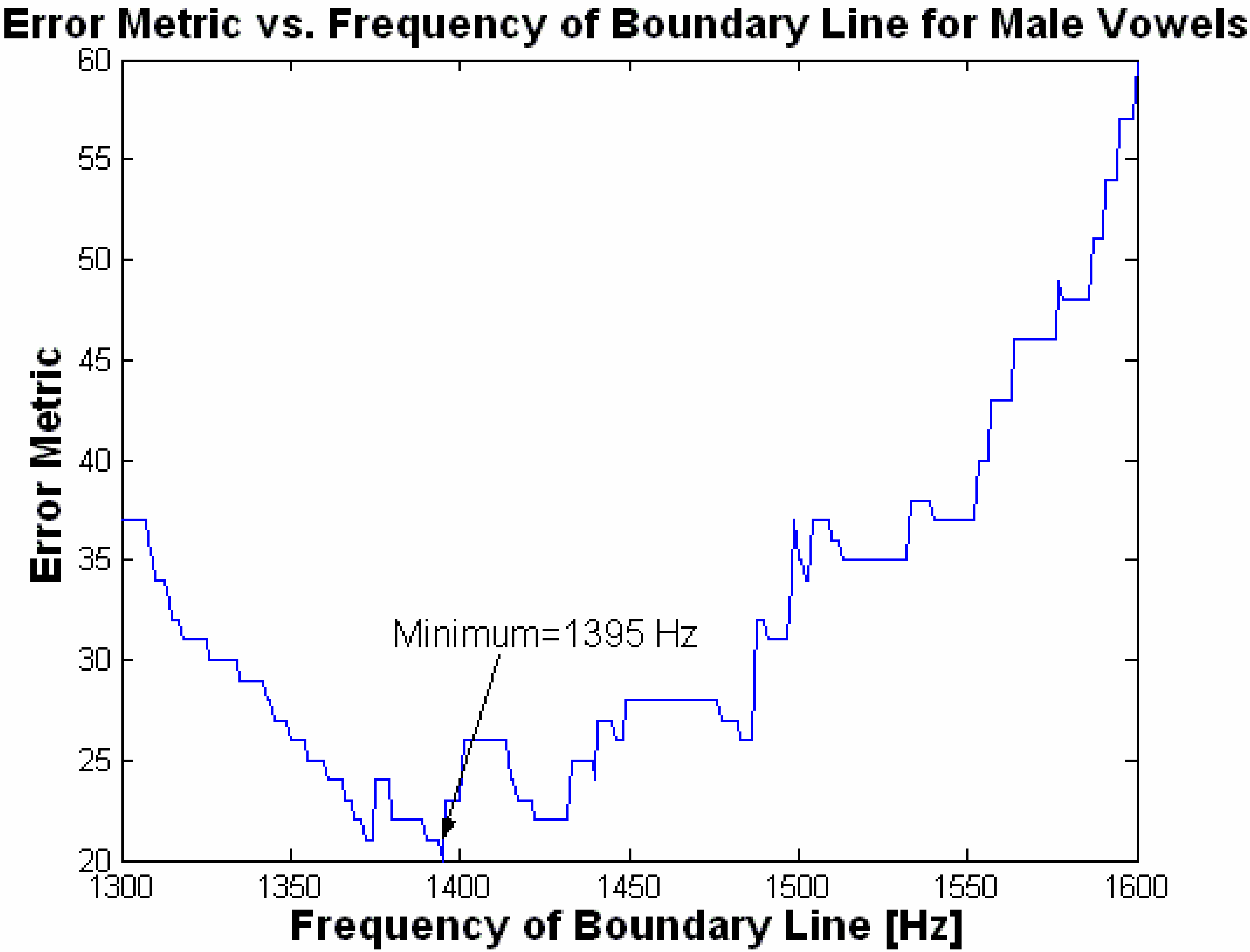
Cross-Linguistic Data

- So far so good, but maybe this is a pattern of American English vowels.
- Anecdotally, British, Polish measurements also follow pattern.
- Can look at how cross-linguistic vowel formant data patterns.
- 44 male, 18 female surveys, >3 speakers.
- 9 back, 7 central, 9 front vowels, different qualities (short, long, breathy, nasal, laryngealized).

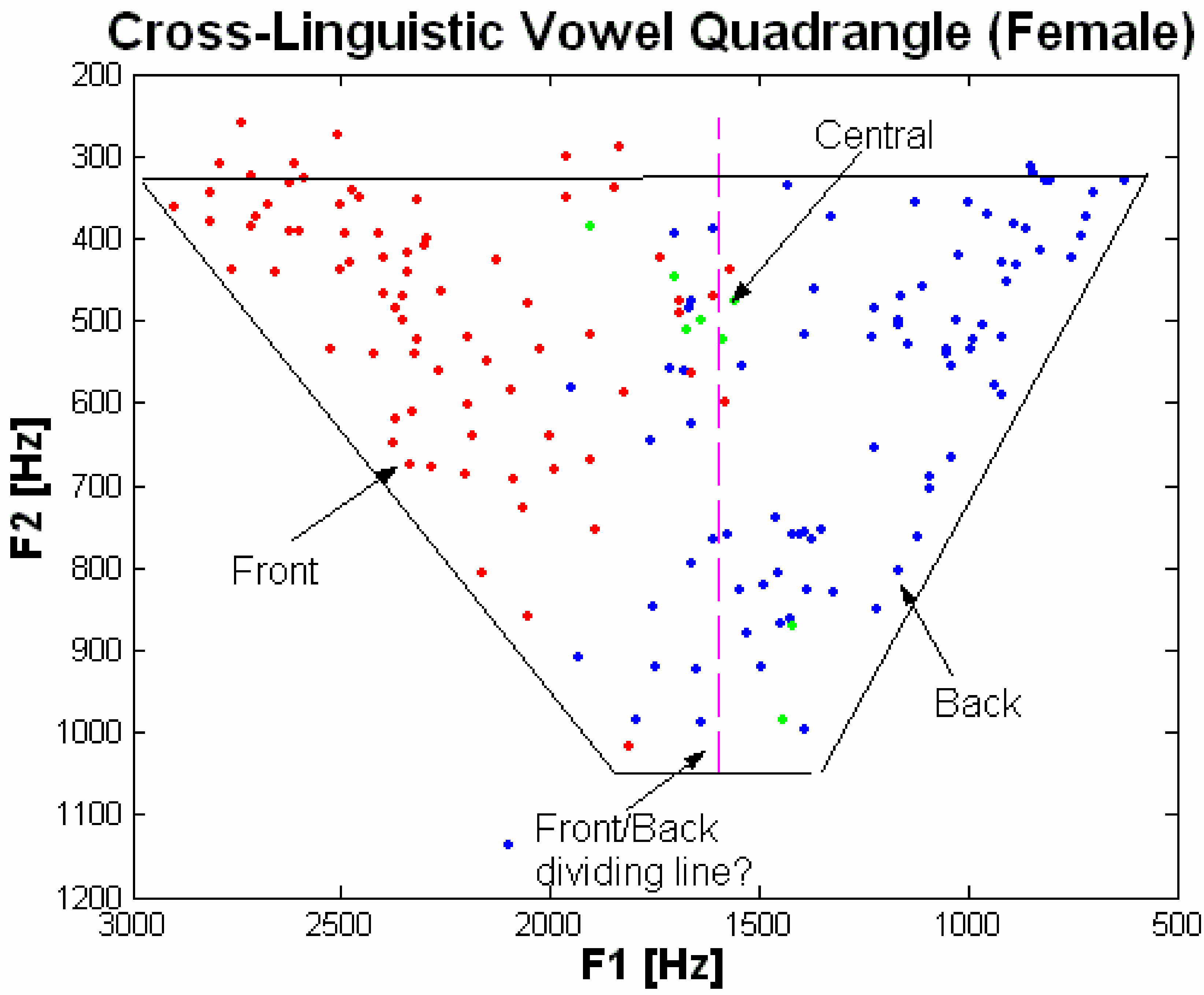
Cross-Linguistic Vowel Quadrangle (Male)



- Relatively sharp front/back division.
- To find where, vary the boundary line frequency, plot the error metric.
- Error metric=(# of back vowels>freq) + (3*# of front vowels<freq).
- Somewhat arbitrary – front vowels must “count” more because back vowels tend to front (much more common than backing diachronically), more lax vowels (less peripheral) are back.


Error Metric vs. Frequency of Boundary Line for Male Vowels
Error Metric
Frequency of Boundary Line [Hz]
Minimum=1395 Hz
60
55
50
45
40
35
30
25
20
1300
1350
1400
1450
1500
1550
1600

- Find boundary line ~1395 Hz, agrees with subglottal data averaged with other studies (1355± 56 Hz).
- 4.7% of front/back vowels on “wrong” side.
- 20 central vowels divided 13/7 by line at 1395 Hz.
- Strong tendency towards hypothesis, same for female data?


Cross-Linguistic Vowel Quadrangle (Female)
F2 [Hz]
F1 [Hz]
200
300
400
500
600
700
800
900
1000
1100
1200
3000
2500
2000
1500
1000
500
Central
Front
Back
Front/Back
dividing line?

- Find boundary line ~1555 Hz, agrees with subglottal data averaged with other studies (1518±104Hz)
- 9.3% of front/back vowels on “wrong” side.
- 8 Central vowels divided 6/2 by line at 1555 Hz.

Cross-Linguistic Results

- Observe dividing effect for male and female data, stronger for males.
- Hard to explain location of boundary line otherwise – even if ~halfway across quadrangle, not true in Barks.
- Still anecdotal – shaky method, few speakers in some studies, bias towards Germanic/IE languages, general unreliability of formant measurements.

Theoretical Implications

- Some support for central vowels being unspecified for [back].
- Another possible reason for why only 3 horizontal classes, versus 5 vertical ones?
- Dispersion theories of vowel space structure: Lijencrants & Lindblom 1972 & passim (“Adaptive Dispersion”), Flemming 1995 & passim (“Dispersion Theory”) in OT.
- Maximize distance between vowels, minimize effort, maximize number of contrasts.

- Both theories take frequency-phoneme map for granted.
- AccF2 may help define this map.
- No reference to features in either theory, but vowel spaces are formed by change acting on features.
- Both theories assume a relatively homogeneous space of possible vowels.
- But some speakers have unstable regions which repel possible vowels with F2 near AccF2.
- Need dispersion attributes+quantal attributes?

Conclusion

- Possible quantal features can be modeled, tested at several levels.
- Hypothesis generally supported at all levels => AccF2 may give front/back distinction.
- Possibly a quantal feature, certainly a phonetic tendency.
- Should be enough that it’s generally true – many aspects of languages are biases, not universals.
- Many thanks to Professor Ken Stevens and members of the Speech Communication Group.
- Work supported by NIH Grant DC00075.

References

Languages used in the formant survey, references available from author MS.

J. van den Berg, “An electrical analogue of the trachea, lungs and tissues.” *Acta Physiol. Pharmacol. Neerlandica* **9**: 1-24 (1960).

H. Cheyne, *Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck*. Ph.D. dissertation, MIT (1993).

B. Cranen & L. Boves, “On subglottal formant analysis.” *JASA* **81**: 734-746 (1987).

E. Flemming, *Auditory Representations in Phonology*. Ph.D. dissertation, UCLA (1995).

K. Ishizaka et. al., “Input acoustic impedance measurement of the subglottal system.” *JASA* **60**: 190-197 (1976).

D. Klatt & L. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers.” *JASA* **87**: 820-857 (1990).

J. Lijencrants & B. Lindblom, “Numerical simulations of vowel quality systems: The role of perceptual contrasts.” *Language* **48**, 839-862 (1972).

K.N. Stevens et. al., “A miniature accelerometer for detecting glottal waveforms and nasalization.” *J. Speech and Hearing Research* **18**: 594-599 (1975).

K.N. Stevens, “On the quantal nature of speech.” *J. of Phonetics* **17**: 3-46 (1989).