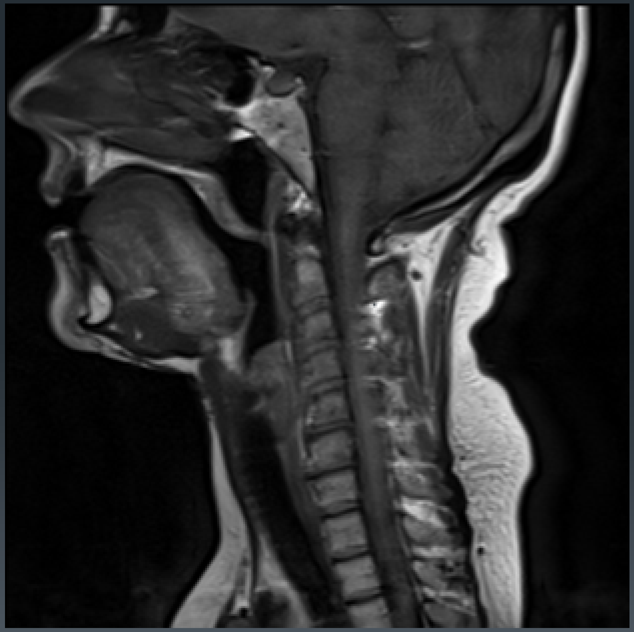
A descriptive approach to the measurement of nasalization



Ryan Shosted
University of Illinois at Urbana-Champaign
NWAV 41, Bloomington IN
Towards Best Practices in Sociophonetics

Outline

- General problems
- Perception
- Production
  - Articulation
  - Aerodynamics
  - Acoustics
- Acoustic measurement
  - Time domain
  - Frequency domain
- Aerodynamic measurement: A beginner's guide

General problems: What is nasalization?

- Nasalization is at once an articulatory, aerodynamic, acoustic, and (ultimately) perceptual phenomenon.
- We can, for example, estimate the degree of VPO that it takes for a sound to be perceived as nasalized (>10mm2, Warren 1993 et al.) or to result in certain acoustic properties.
- Certain acoustic properties can be shown to result in a percept of nasalization (NB: they may not all be strictly associated with VPO)
  - Related to VPO: F-frequency, F-amplitude, Anti-F, SP amplitude
  - Unrelated to VPO: Length, phonation type, f0(?)

General problems: Why measure different aspects of nasalization?

- Perceptual
  - In controlled experiments, we can determine which acoustic characteristics are most important to a percept of nasality;
  - We have discovered that some acoustic characteristics unrelated to VPO are associated with nasalization (e.g., length, quality)
- Acoustics: Acoustic measures are indirect indications of VPO; also, *it’s what we hear* (cf. Ohala 1996)
- Aerodynamics: An indirect measure of VPO
- Articulation: VPO is the “unambiguous” cause of nasalization (cf. point 2 about perception)

General problems: Practical problems in measurement

- **Degree of nasality**: How 'nasal' is it?
- **Temporal characteristics of nasalization**: When does nasalization start/end?
- **Vocal tract configuration**: What is the place of articulation of nasal consonants / the area function of nasal vowels (including complex nasal geometry + sinuses)?
- Which production mechanism should be measured? Acoustics, aerodynamics, or articulation?

General problems: Applied problems (for linguists)

- Phonemic vs. allophonic nasalization: Are there phonetic differences? How can I tell when I am observing one or the other?
- Nasal spreading / harmony / coarticulation / coproduction: Distance, directionality, and degree of nasalization caused by some ‘trigger’
- What gives rise to a nasal percept? Is it best measured in the acoustic, aerodynamic, or articulatory domain?

Perception: What gives rise to a nasal percept?

- Time domain
  - Length
  - Reduced sound pressure amplitude
- Frequency domain
  - Reduced F1 amplitude (Delattre 1954; House and Stevens 1956)
  - Pole-zero (nasal formant-antiformant) pair around 1 kHz (Hawkins and Stevens 1985)
  - Increased bandwidth, particularly of F1 (Huffman 1990)
  - Additional peak 250 – 450 (pnasal sinuses) (Hattori et al. 1958)
  - F1-prime (perceptual merger of F1 and nasal formant)

Production: Articulation

- Why? Provides direct evidence of one or more speech gestures
  - Occurrence: VPO / velum lowering; oral closure; modifications to non-VP speech articulators
  - Duration: Anticipatory or perseverative velum lowering / VPO; temporal extent of non-VP gestures
  - Dynamic aspects: Increasing / decreasing VPO; velocity of velum lowering / raising
- Special concerns: The relation between the percept of nasality and an articulatory variable like VPO is non-linear; e.g., VPO is not always necessary to generate a nasal percept.

Production: Aerodynamics

- Why? Provides **relatively** unambiguous evidence of one or more speech gestures
  - Occurrence: VPO / velum lowering; oral closure
  - Duration: Extent and direction of ‘nasalization’
  - Dynamic aspects: Changes in ‘nasalization’ over time
- Special concerns: Nasal flow ≠ VPO
  - Nasal flow (incl. proportional nasal flow) is affected by oral impedance: High vowels have more nasal flow but are not necessarily more ‘nasal’ in any other sense
  - Flow is generally affected by subglottal pressure (loudness, f0(?))

Production: Acoustics

- Why? It is what listeners hear – we know nasalization **must be** encoded in the acoustics
- Why not? Provides perhaps the *most* ambiguous evidence of one or more speech gestures
- Special concerns
  - Time domain: Amplitude gradient (e.g. boundary between nasal vowel and nasal consonant)
  - Frequency domain: Evidence of anti-formants is deductive (based on absence of energy); Harmonics are difficult to identify reliably; LPC does not account for anti-formants – so whither the formant (frequencies and amplitudes)?; Many-to-one relationship between articulation and acoustics (e.g., is *f* attributable to VPO or tongue position?)

Production: Which aspect of production should be measured?

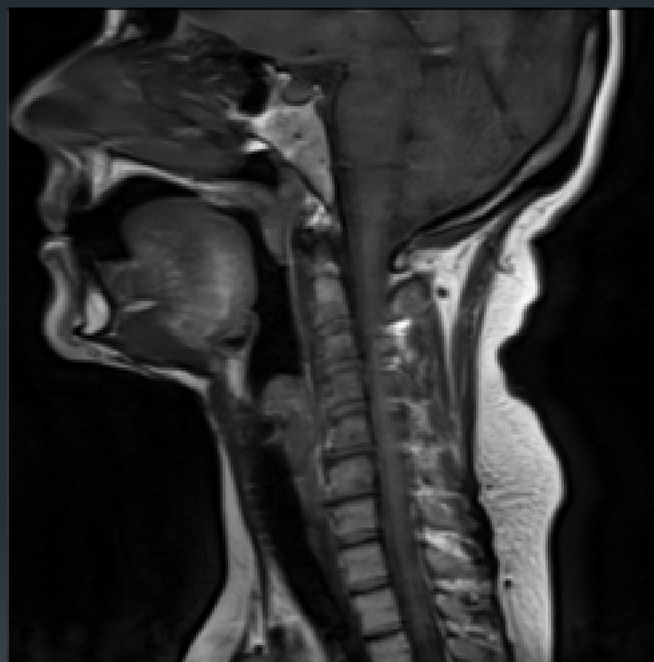
- A mixed approach will be the most enlightening
- Remember the “practical problems for measurement” introduced earlier?
  - Degree of nasality: Aerodynamics and/or articulation + acoustics
  - Temporal characteristics of nasality: Aerodynamics and/or articulation + acoustics
  - Vocal tract configuration: Aerodynamics and/or articulation + acoustics

Production: What if I just measure the acoustics?

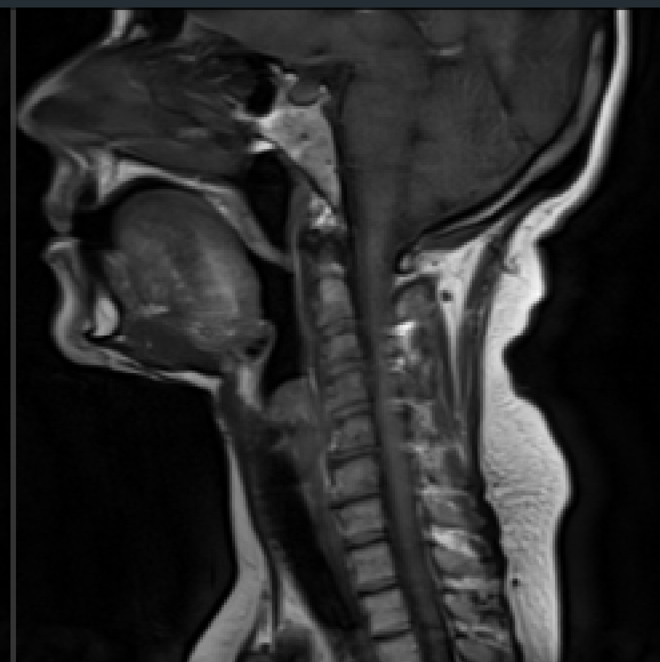
- Degree of nasality:
  - No consensus on how to measure the degree of nasality (Pruthi and Espy-Wilson 2007 use nine different measures!)
  - For formant-amplitude schemes, it is not always possible to identify the relevant formants **because** of what nasalization does to the signal
  - Identifying peaks vs. measuring spectral tilt
- Temporal characteristics
  - High variability in choosing boundaries, e.g., between nasal vowels and nasal consonants
  - Is degree of nasality increasing over time? See above.
- Vocal tract configuration
  - For subtle effects (e.g., changes in oro-pharyngeal vowel quality) it's impossible to separate nasalization from the oro-pharyngeal area function (NB: the lowered velum affects the area function, too)
  - LPC is not designed to handle anti-formants so estimates of formant frequencies and amplitudes may be highly variable

Production: When acoustics have required supplement

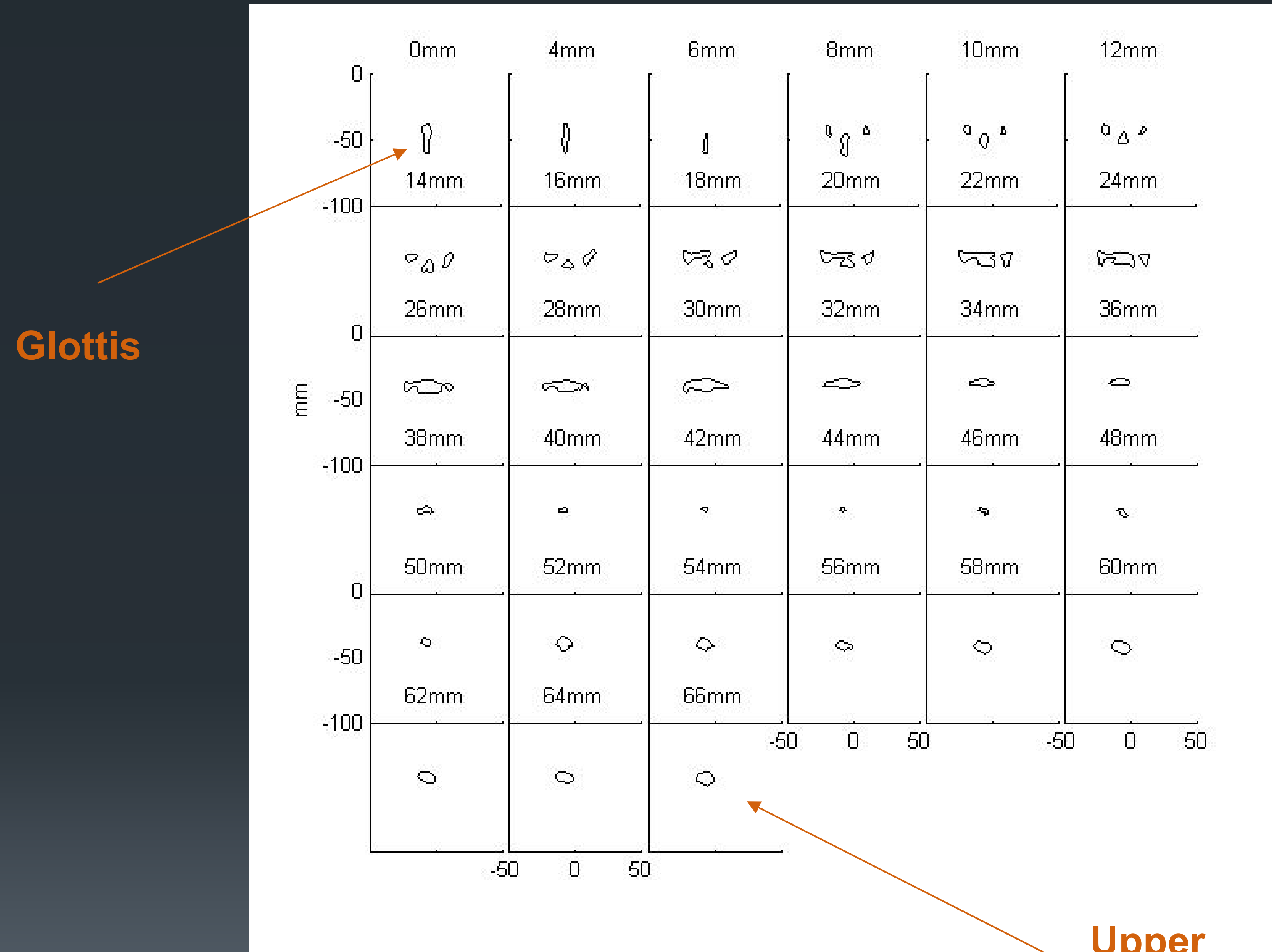
- /nts/ vs. /ns/ sequences: Difficulties in identifying nasal closure vs. vowel nasalization (Shosted 2010)
- Nasal codas after nasal vowels (Shosted 2006, 2011)
- Nasal / nasalized vowel quality
  - Brazilian Portuguese: is the nasalized low vowel perceptually raised as a consequence of nasalization, tongue position, or both (Shosted, sub.)?
  - Are chain shifts and mergers of nasal vowels, historically attested in French, based on nasalization, tongue position, or both (Carignan 2011)?
  - Do articulatory characteristics support the maintenance of oral / nasal and/or nasal / nasal contrasts in Hindi (Shosted et al. 2011)?
  - Do speakers compensate for allophonic nasalization (Arai 200X)
- What is the temporal extent of nasalization (Devlaux et al. 2008)?
- Speaker characteristics: Does nasal cavity size matter in the production of nasal vowels (Engwall et al. 200X)



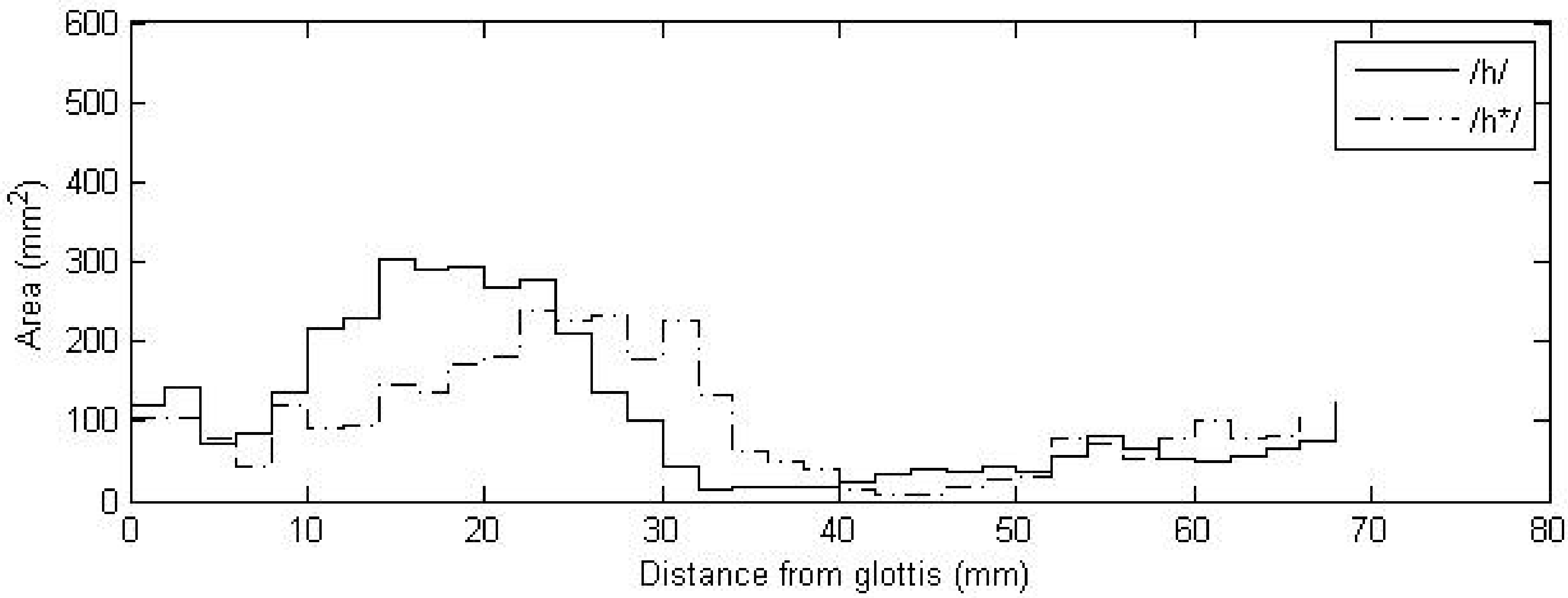
Oral /u/



Nasal /u~/



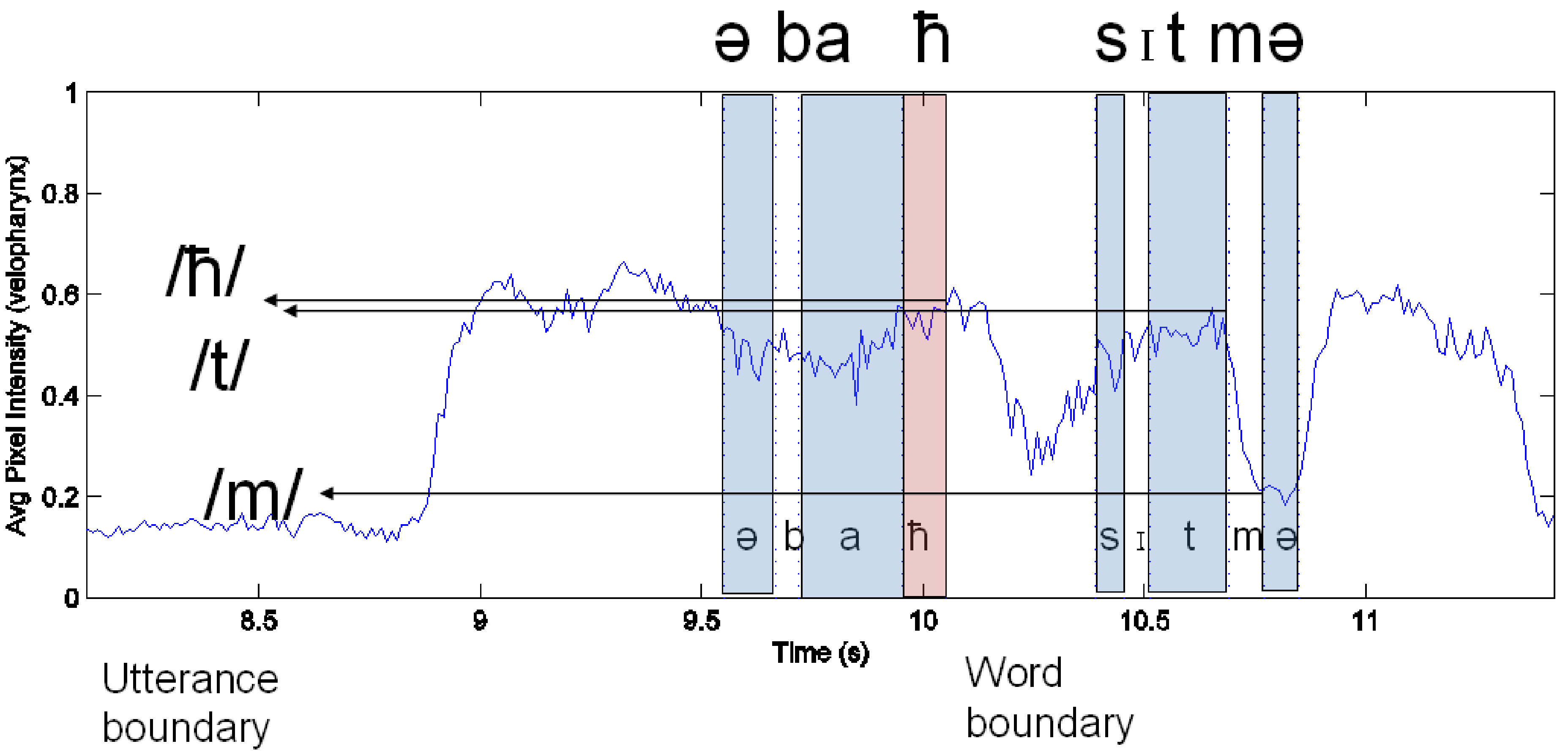
Arabic pharyngeal fricative [ħ]



/h/
/h*/
Area (mm2)
0
100
200
300
400
500
600
0
10
20
30
40
50
60
70
80
Distance from glottis (mm)

Real-time MR

- 86 frames/s
- 2.2 x 2.2 x 6 mm voxel
- Reconstructed using partially separable functions (Liang 2007)
- *Click here: iktubu rabah sit marrat*
- *Go to web: digo aipim agora*



pace Zellou (2012)

Velopharyngeal port

Fig. 2. Midsagittal image that provides a reference for proper orientation of the coronal oblique slice (at left). Sample of four coronal oblique frames obtained during the production of the nasal vowel [6˜], in the word [6˜p6] (at right).

Teixeira, Oliveira, Martins, Ferreira, Silva, and Shosted (2012: 309)

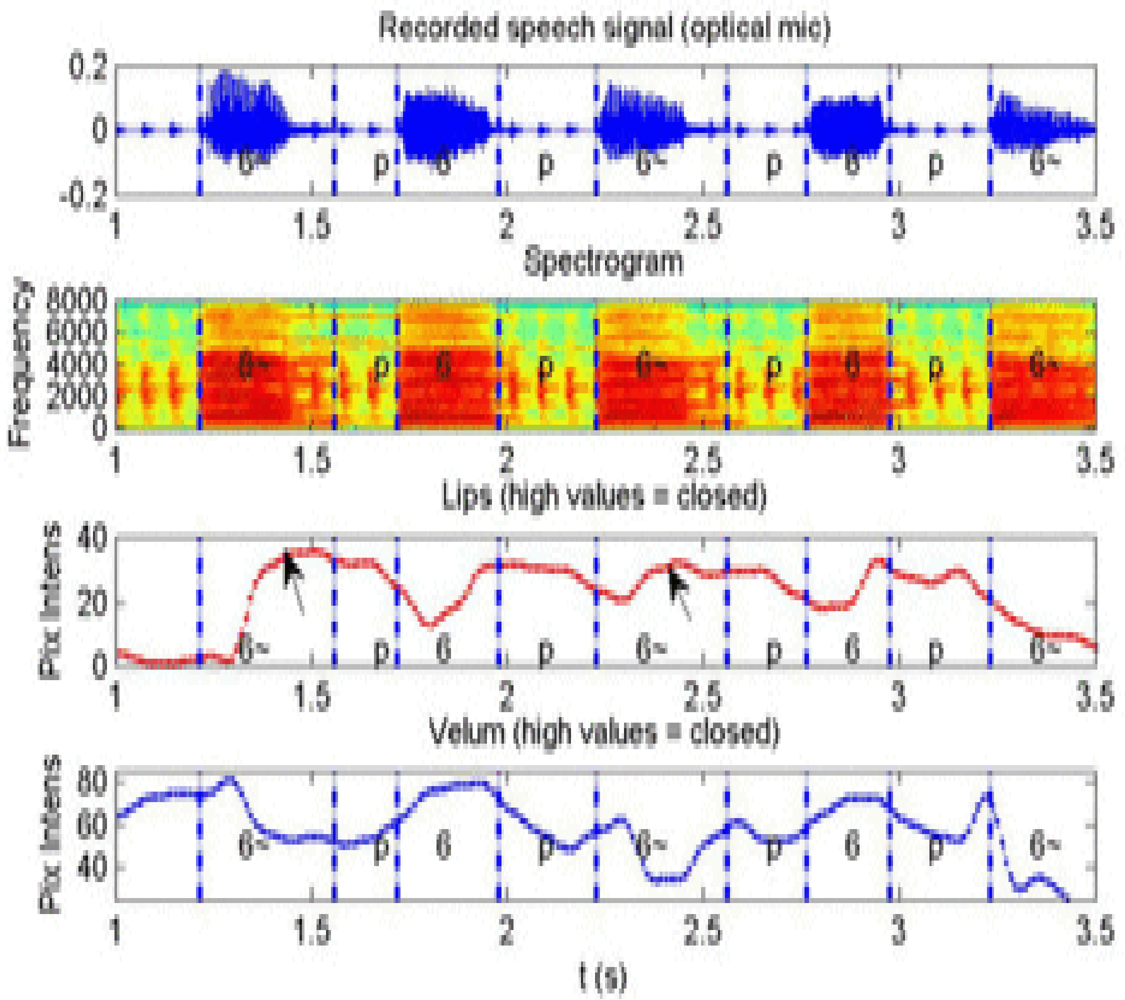
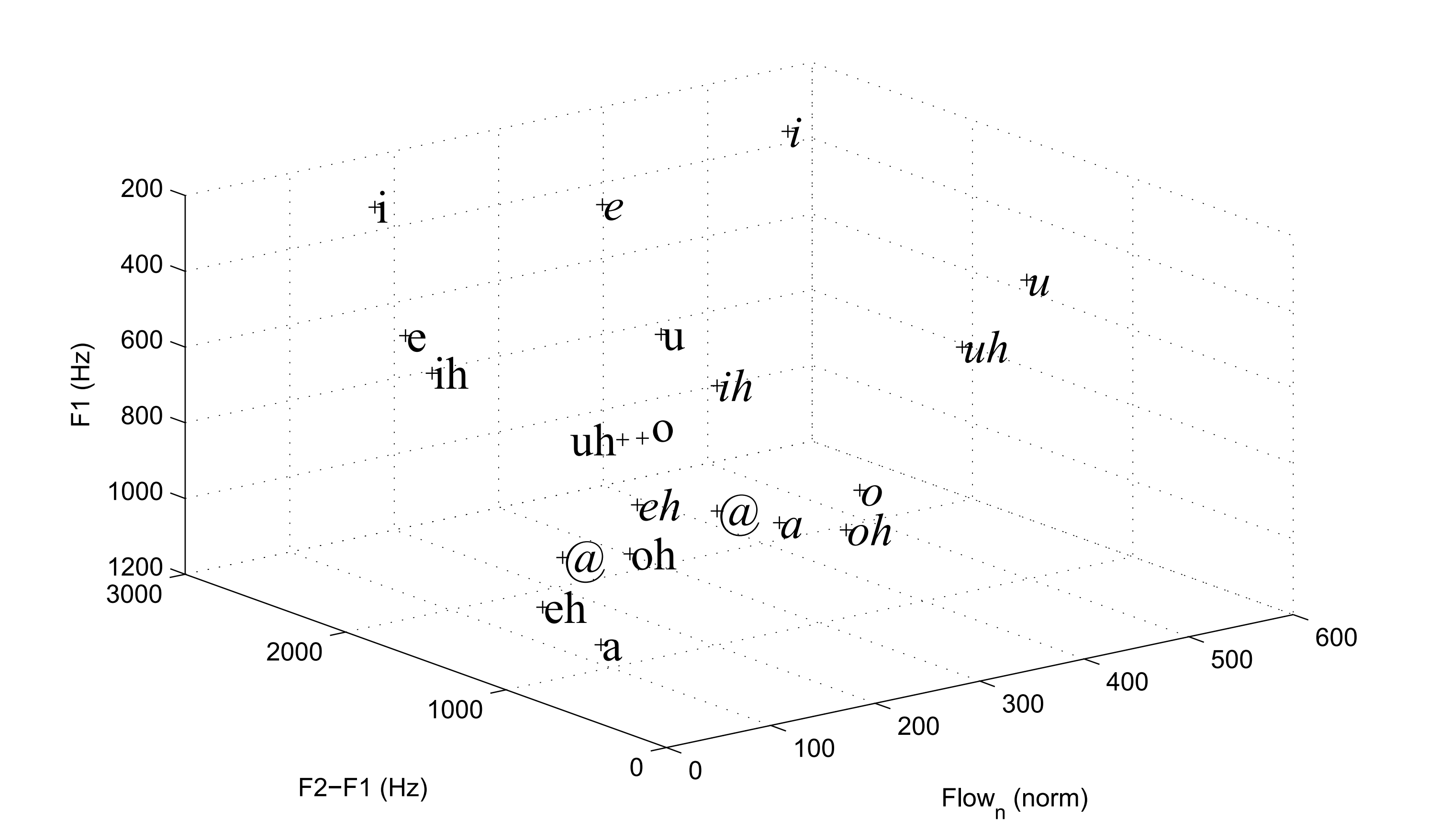
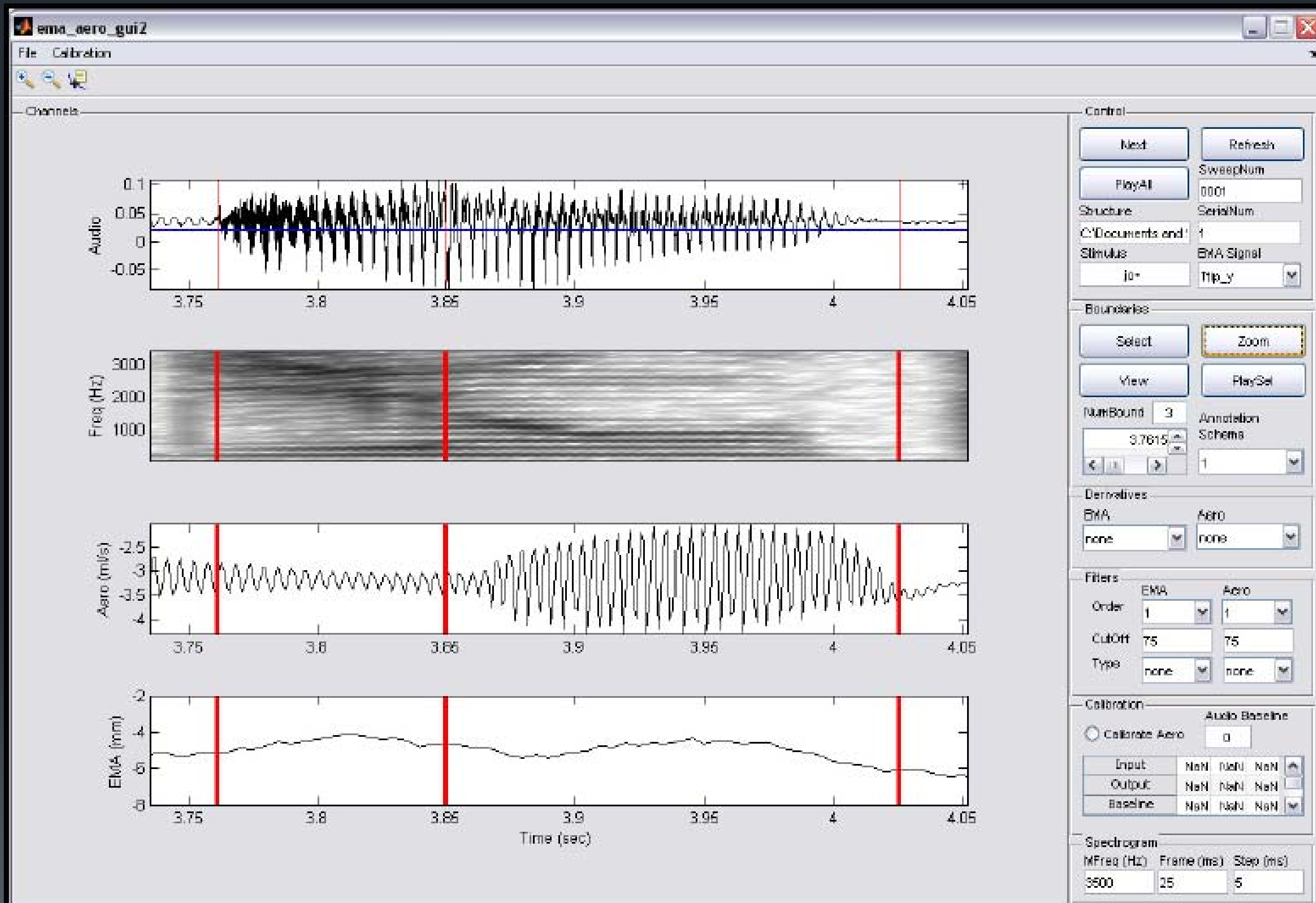


Fig. 6. Recorded speech signal from fiberoptic microphone (top), changes in intensity over time, in the selected ROI's: lips (3rd row) and velum (bottom). The speech segment presented corresponds to the production of [6˜p6], [p6˜p6], [p6˜](2.5 s of speech).

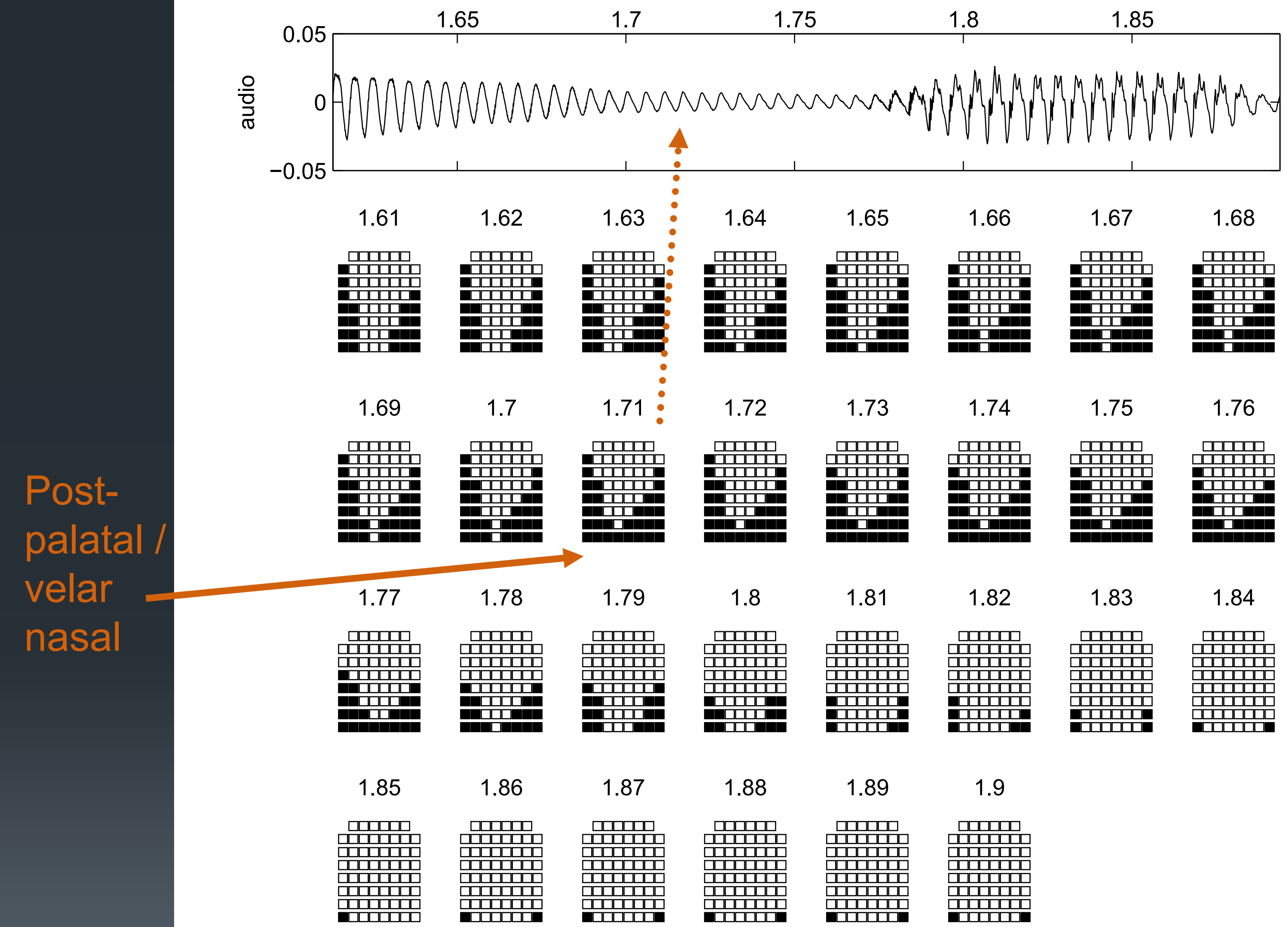
Teixeira, Oliveira, Martins, Ferreira, Silva, and Shosted (2012: 312)



F1 (Hz)
200
400
600
800
1000
1200
3000
2000
1000
0
F2−F1 (Hz)
0
100
200
300
400
500
600
Flow$_n$ (norm)
i
e
ih
u
uh
o
eh
@
oh
a
i
e
u
uh
ih
o
eh
@
a
oh



Hindi /o~/ (Shosted et al. 2012)



Brazilian Portuguese <im#a> (Shosted 2011)

Acoustic measurement: Time domain {Length}

- The perception of vowel nasalization is favored by increasing vowel duration (Lintz and Sherman 1961; Cagliari 1977; Delattre and Monnot 1981; Whalen and Beddor 1989)
- Does this mean nasalized vowels are longer than their non-nasal counterparts? To test the hypothesis:
  - Look for a drop in the amplitude of the acoustic waveform
  - Consider setting a threshold for this drop to make the result both falsifiable and replicable
  - Ideally, such a threshold should be set based on aerodynamic or other articulatory evidence

Acoustic measurement: Time domain {SP amplitude}

- There is a general reduction in sound pressure amplitude associated with nasalization (due to greater sound absorption by the nasal turbinates, i.e. the considerable surface area of the nasal cavity = 3.5 x greater, Bjuggren and Fant 1964)
- Ideally, thresholds should be set: wording like “sudden drop in amplitude” or “dramatic drop in amplitude” are obviously difficult to replicate and falsify
- Possible solutions:
  - Inter-rater reliability (note that this is not traditional IRR, since the choice of a boundary is not categorical, but gradient)
  - Automatic segmentation, e.g., the Penn Phonetics Lab Forced Aligner (Yuan and Lieberman 2008)

Acoustic measurement: Frequency domain {F1}

- For vowels, the frequency of F1 is generally **increased** by nasalization, i.e., vowel quality is lowered during nasalization (Fujimura and Lindqvist 1971)
- This is due to the appearance of a nasal formant **above F1** (for non-low vowels) which spreads the distribution of energy **higher**
- For low vowels (like /a/) which already have a high F1, the nasal formant is **below F1**, so the energy is distributed **lower** than usual (**lowering F1** and thereby **raising** the vowel) (Kluender et al. 1990) – this tends to be the case with heavy nasalization

Acoustic measurement: Frequency domain {N1}

- First nasal formant (N1) appears below F1 for low vowels (at least when heavily nasalized) and above F1 for non-low vowels (this is the main reason for the centralization of nasal vowels often referred to, e.g., in Beddor 1983)
- The presence of N1 is best deduced from comparison with an oral version of a vowel
  - This is great for paradigmatic comparison of, e.g., /a/ and /a~/
  - For syntagmatic comparison of the oral and nasalized part of a vowel, how can you tell when N1 has appeared?
  - One solution is to work backwards by identifying the nasal formant, e.g., the end of a vowel in a VN sequence and then measuring its amplitude farther away from the nasal ‘trigger’

Acoustic measurement: Frequency domain {F-amplitude schemes}

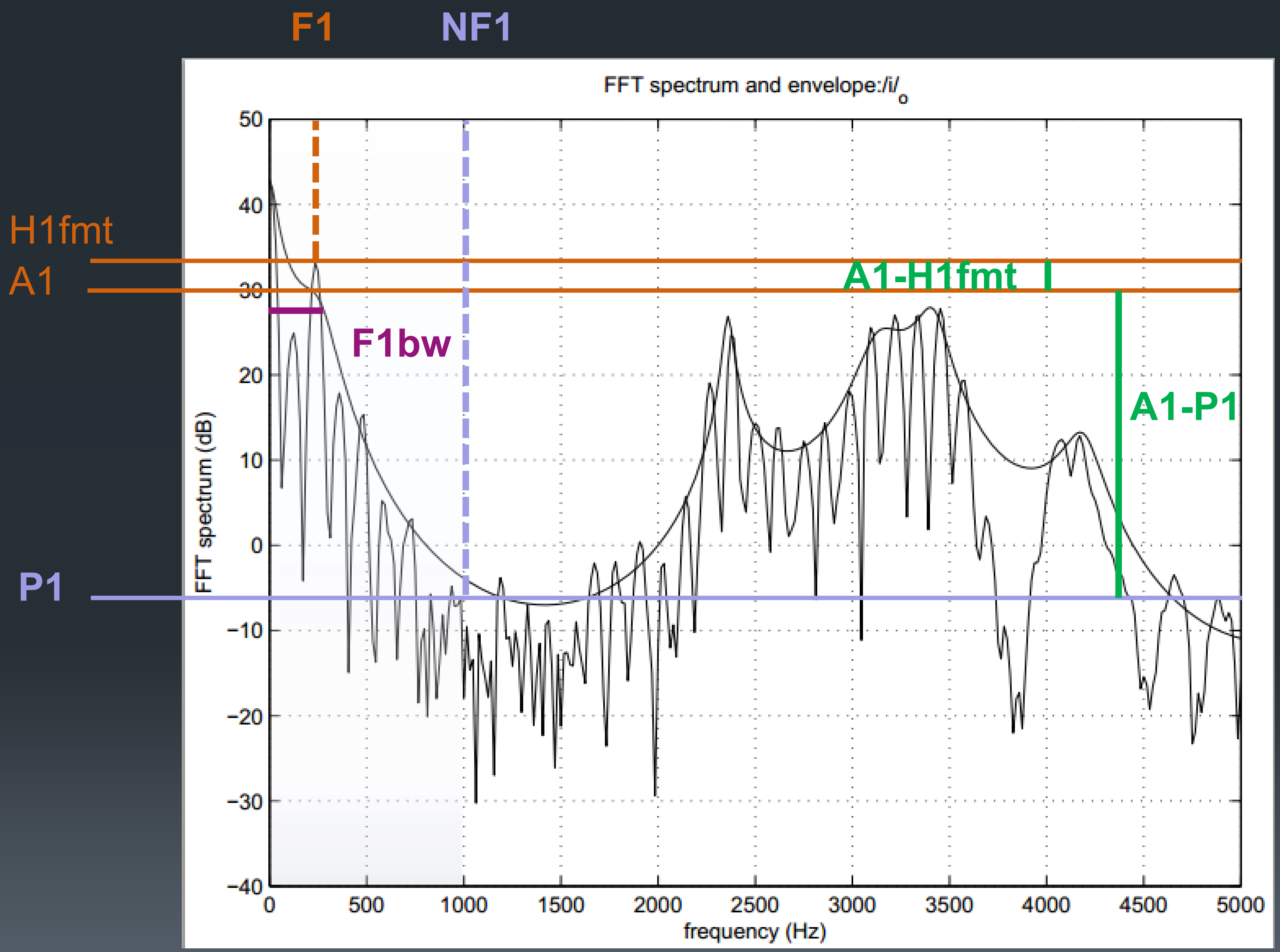
- A1 − H1 (Huffman 1990)
  - Amplitude of F1 minus amplitude of first harmonic (H1)
- N2 − N1 (Maeda 1993)
  - Amplitude of N2 minus N1: N1 and N2 may be the first formant peak, the nasal formant peak, or the second formant peak, depending on which two of the three peaks are stronger.
- A1 − P0 (Chen 1997)
  - Amplitude of F1 minus amplitude of first nasal formant (P0) "often below F1"
- A1 − P1 (Chen 1997)
  - Amplitude of F1 minus amplitude of second nasal formant (P1) between F1 and F2

Acoustic measurement: frequency domain {F-amplitude schemes} (2)

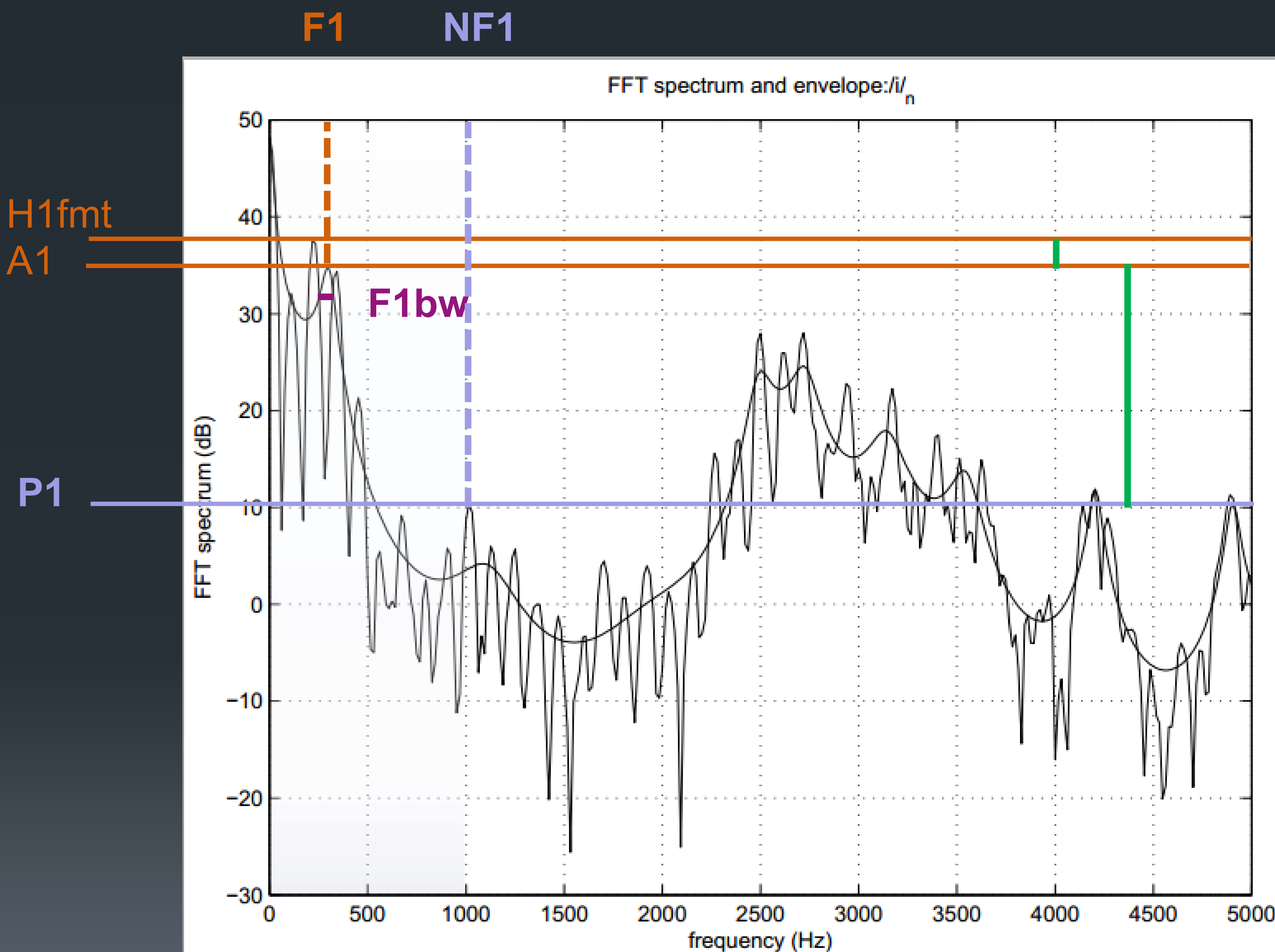
- Which one to choose?
- Huffman (1990) found a good match between the perceived degree of nasalization and A1 − H1 for [i] and [I] but not for [ae]
- Maeda (1993) found a good match between the perceived degree of nasalization and N2 − N 1 for [i] and [o] but not for [u]
- P0 can be difficult to distinguish since it often occurs below, i.e., on the skirt of F1 (Chen 1997)
- Does the nasal-formant really stay in the same position?

Acoustic measurement: Frequency domain {Much, much more}

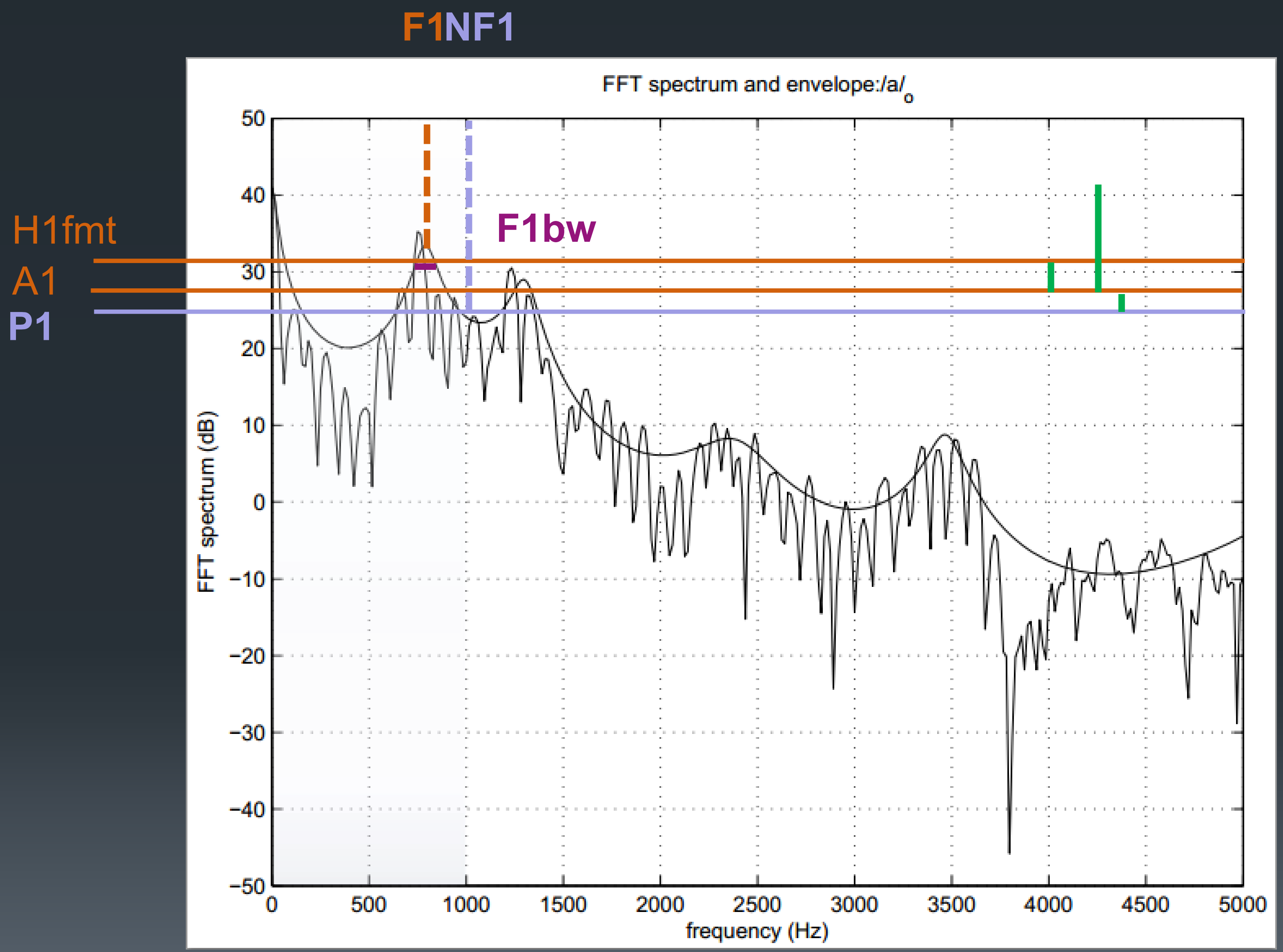
- In building a support vector machine to automatically classify Hindi nasal and oral vowels, Pruthi and Espy-Wilson (2007) use the following (additional) frequency-domain measures:
  - teF1 = correlation of Teager energy profile (Cairns et al. 1996) passed through a narrowband (100 Hz bw) and a wideband filter (1 kHz bw) around F1 [lowers for nasal]
  - nPeaks40dB = Number of formants with amplitudes within 40 dB of the formant with the highest amplitude [raises for nasal]
  - A1-H1max800 = Amplitude of oral F1 minus amplitude of H1 [lowers for nasal]
  - A1-H1fmt = Amplitude of oral F1 minus amplitude of harmonic closest to F1 [raises for nasal]
  - F1BW = bandwidth of F1 [raises for nasal]
  - std0-1K = standard deviation of center of spectral mass 0--1kHz [raises for nasal] (Glass and Zue 1985)



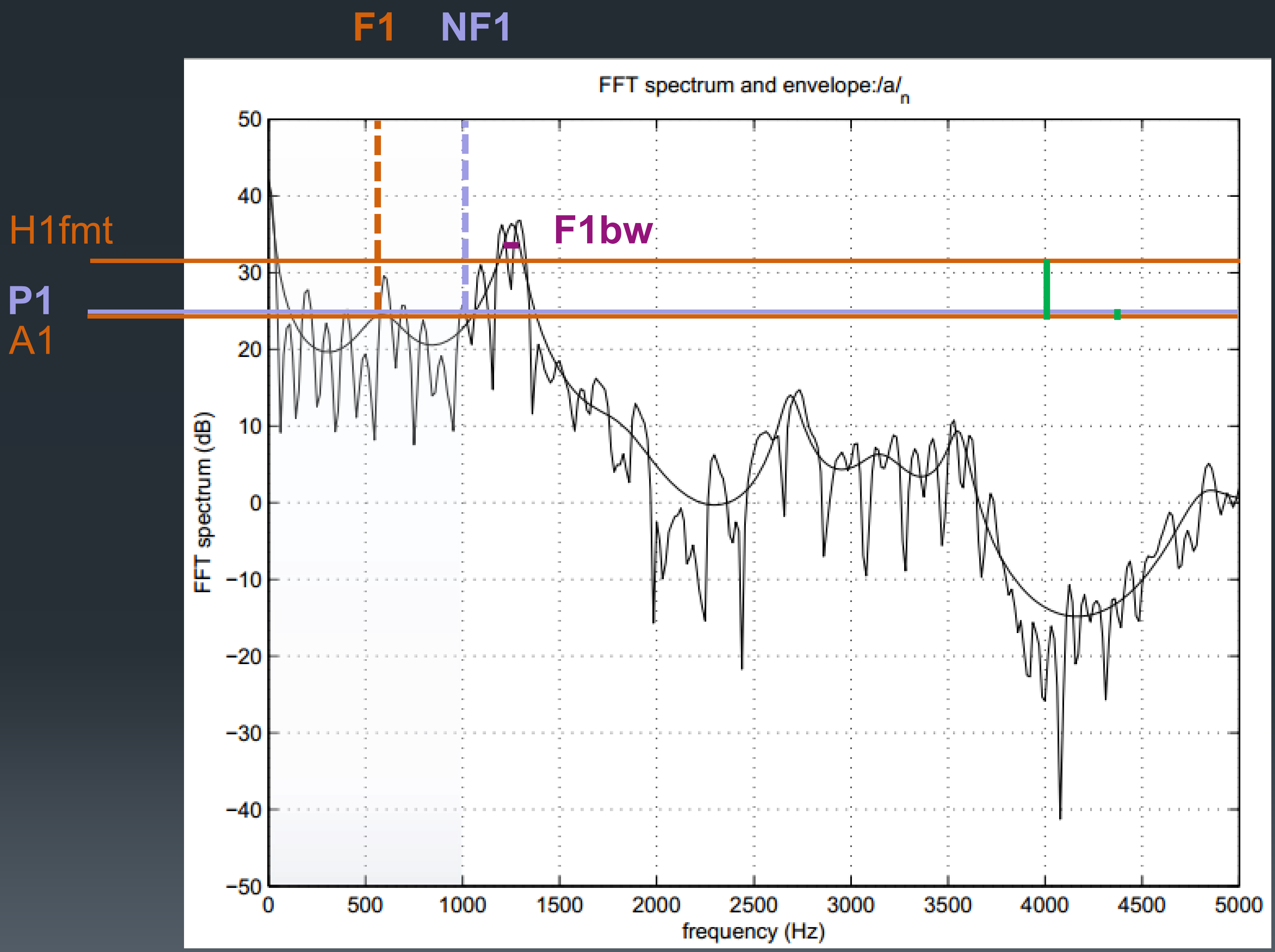
Shosted, Carignan and Rong (2012)



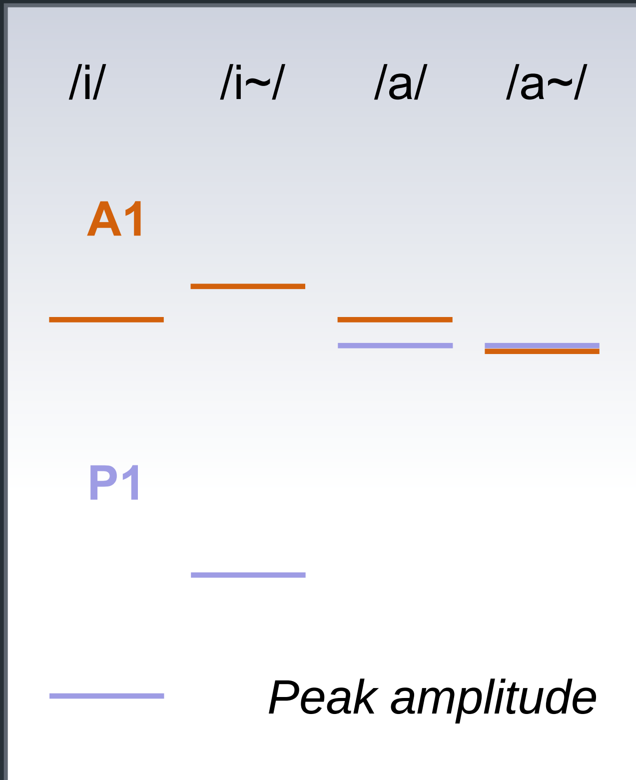
Shosted, Carignan and Rong (2012)

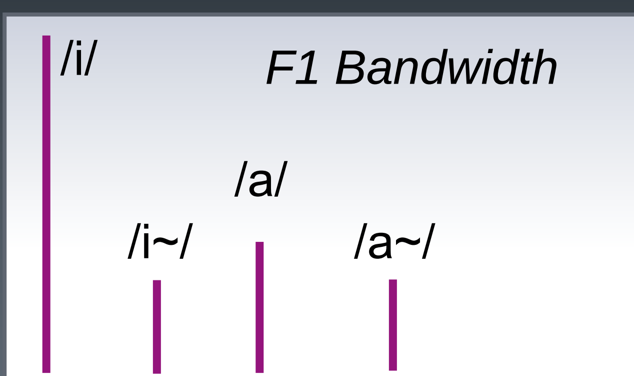


Shosted, Carignan and Rong (2012)



Shosted, Carignan and Rong (2012)





Predictions:
Nasal A1
=lower ☺

Nasal P1
=higher ☺

Nasal (A1-P1)
=Lower ☺

Nasal (A1-H1fmt)
=Lower ☺

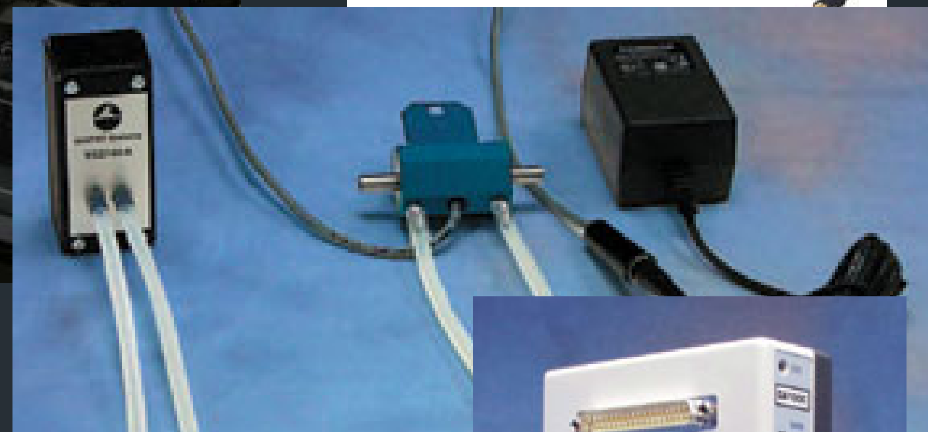
Nasal bandwidth
=Higher ☹

Suggestions for dealing with nasal vowel quality

- 16th- and 30th- order LP filters designed using Matlab's FFT function
- Peaks in the LP filter are detected autoatically
- Figures of each 1024-point FFT spectrum with an LPC overlay were generated and the F2 x F1 vowel space was plotted for each vowel.
- More reasonable peaks are picked by hand if they diverge greatly from the vowel's F2 x F1 profile
- For NVs, the nasal formant is first identified, followed by F1 and F2 (must be comparative: Nasal vs. Oral)
- How else can we map oral and nasal vowels into the same space? Can we abandon F1 and F2 when working with nasal vowels?

Aerodynamic measurement: A bgeinner’s guide {Hardware}

- Hardware requirements: Expensive stand-alone systems are NOT required
  - Pressure transducers
  - Pneumotachs (flow)
  - Calibration syringe (flow)
  - Airflow masks – split masks and/or nasal CPAP masks
  - Tubing (pressure)
  - Manometer (pressure)




BIOPAC Systems
TSD160A
BIOPAC Systems
TSD160B
BIOPAC Systems
TSD160C

Aerodynamic measurement: A beginner's guide {Software}

- Special software is NOT required (but it helps)
  - Consider recording audio and nasal flow as stereo in Praat
  - Most computer sound cards will allow for stereo recording
  - (You will need some kind of audio mixer and the correct connections between your transducers and the mixer)
  - Hypothetical setup:
    - Channel 1 / R: Mic (XLR) → sound mixer → computer
    - Channel 2 / L: Pressure transducer (1/8'') → sound mixer (using 1/4'' adapter) → Computer
- What can special software add?
  - Semi-automatic calibration, automatic filtering, real-time display

Conclusions

- Acoustic measures have the advantage of being easy to obtain but relatively difficult to implement and interpret
- Articulatory and aerodynamic measures have the disadvantage of being relatively difficult to obtain but relatively easy to implement and interpret
- What's next?
  - Large scale validation of acoustic measures using aerodynamic and articulatory data
  - More precise imaging of the nasal tract
  - Implementations of standard nasalization measures (acoustic) in an easy-to-use software environment

Thanks!

- Research supported by University of Illinois Research Board
- NSF 1121780: Ryan Shosted, Zsuzsanna Fagyal, Christopher Carignan
- Thanks to Biomedical Imaging Center staff and collaborators Brad Sutton, Zhi-Pei Liang, and students Maojing Fu and Li-Hsin Ning