Human voice phoneme directivity pattern measurements

Brian FG Katz
Fabien Prezat
Christophe d'Alessandro

Perception Située
LIMSI-CNRS
BP 133, F91403 Orsay, France
brian.katz@limsi.fr

Previous major works in Voice Directivity

- H. K. Dunn and D. W. Farnsworth, “Exploration of Pressure Field Around the Human Head During Speech,” JASA v.10(3) (1939).
- James L. Flanagan, “Analog measurements of sound radiation from the mouth,” JASA v.32(12) (1960).
- Chu, W.T.; Warnock, A.C.C., *Detailed Directivity of Sound Fields Around Human Talkers*, NRC-CNRC Report IRC-RR-104 (2002).

Directivity patterns averaged over long spoken phrases (14-40 sec)

- Malte Kob, *Physical modeling of the singing voice.* Dissertation University of Technology Aachen. Logos-Verlag, Berlin (2002).

Directivity patterns for sung *glissando*

- Various studies on the directivity of artificial mouths, dummy heads, etc.

Current study

Interest

- Evaluate human voice directivity patterns in detail

Hypothesis

- Mouth geometry, and therefore the radiation pattern, vary for different phonemes
- For sustainable phonemes, mouth geometry is fixed
- Consider speech directivity variations at the phoneme level

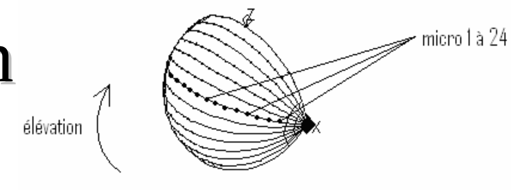
Task

- Measure the directivity of individual phonemes in 3D

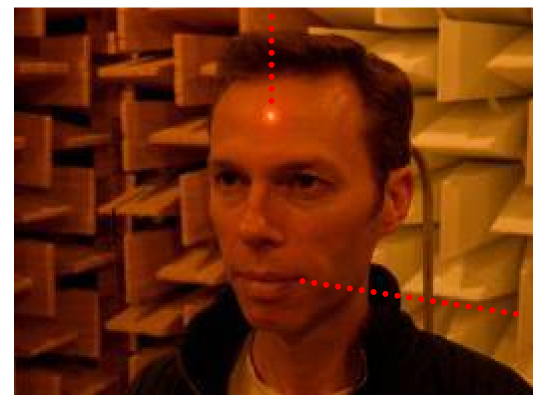
Results

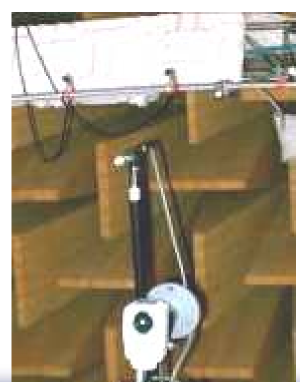
- Preliminary results in 2D (horizontal plane) are presented

Measurement protocol : data acquisition



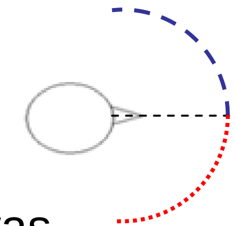
- Sustainable phonemes chosen
  - Vowels: [a] [i] [o]
  - Consonants: nasal [m] [n] & fricative [ch] [s] [f]
- Measurement system (with IRCAM)
  - 180° arc with 24 equally spaced microphones
  - Motorized arc capable of elevations -45° to 90°
  - 2 Reference mics : in-front and head-worn
  - Subject mouth position aligned using fixed laser pointers, head rest assures stable position
  - Video camera at fixed position, and high contrast lip make-up used to help analyze lip forms
  - Audio return (with adjusted reverb) over closed headphones
  - Guitar tuner display under camera used to help subject maintain pitch and level



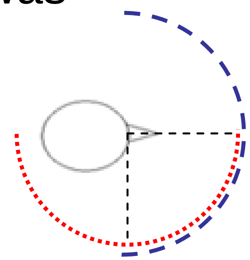


Symmetry / Rotation / Repetition validation

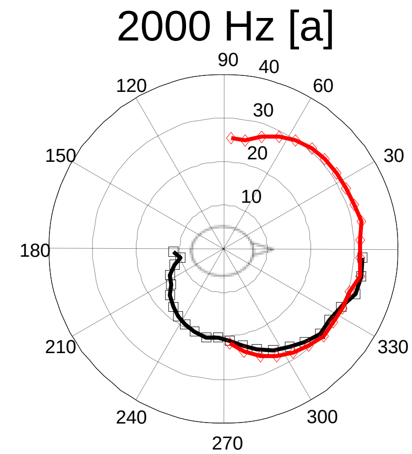
- As our measurement system does not cover the full sphere, or even a full hemisphere, in one pass, either the subject must be rotated or the data must be assumed to be ***symmetrical***.
- Measurements were made using 2 subject orientations : ***forward*** and ***side***.
- Forward symmetry :
  - Mean variation between left & right pattern was < 2 dB

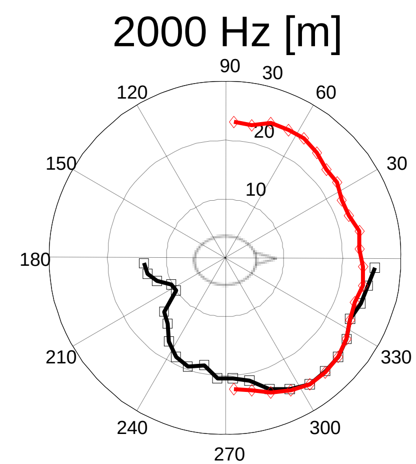


- Rotational symmetry :
  - Mean variation in overlap region between ***forward*** and ***side*** subject orientation was < 2 dB for dominant frequencies for given phonemes

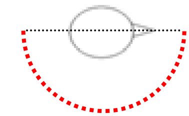


- Repetition variability
  - Mean variation between repetitions was < 1 dB

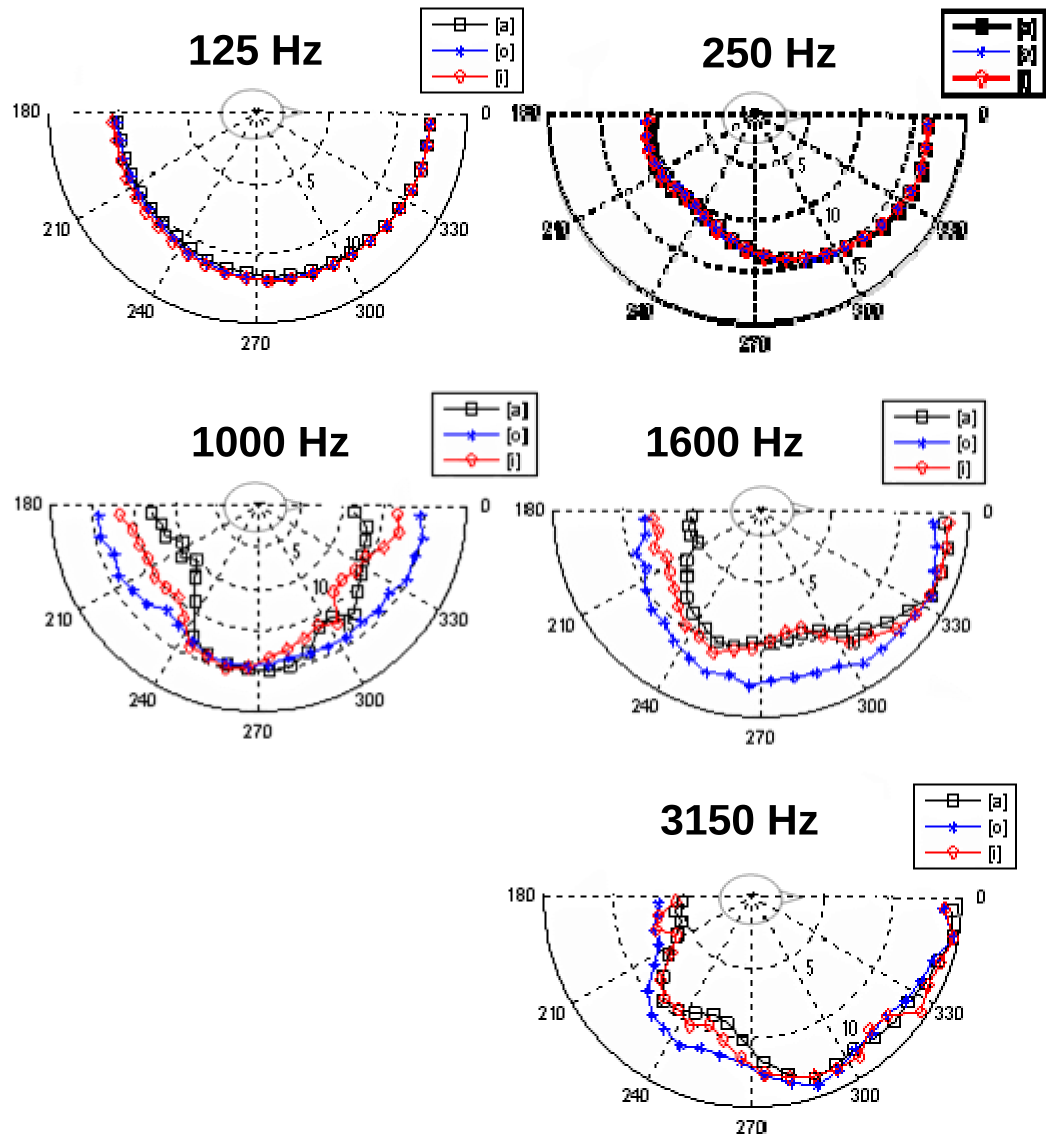




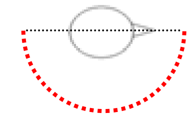
Directivity results : vowels [a][o][i]



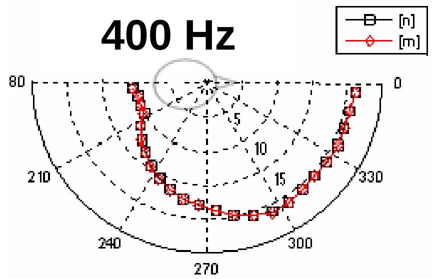
- Data analysis was performed by calculating RMS level in 1/3rd octave bands
- Highly similar at low-frequencies
- Marked variations at mid-frequencies
- High similarity at higher frequencies

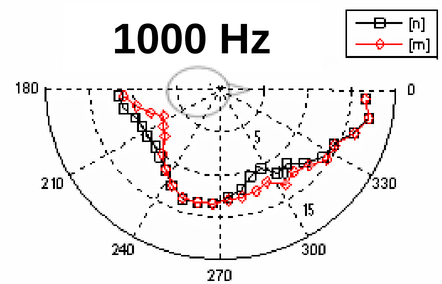


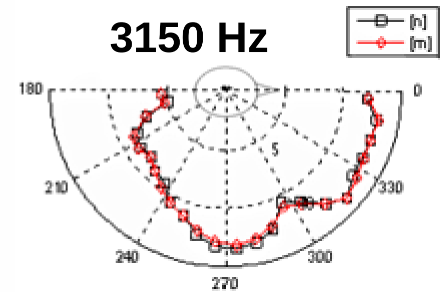
Directivity results : consonants [n][m] & [f][ch][s]



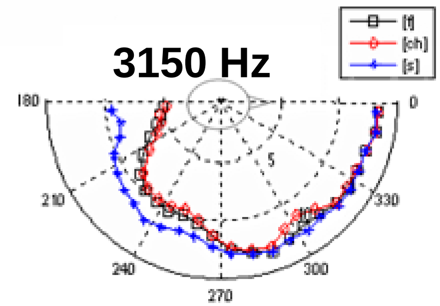
- [n][m] highly similar across frequency range

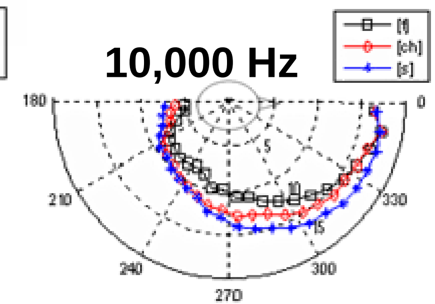




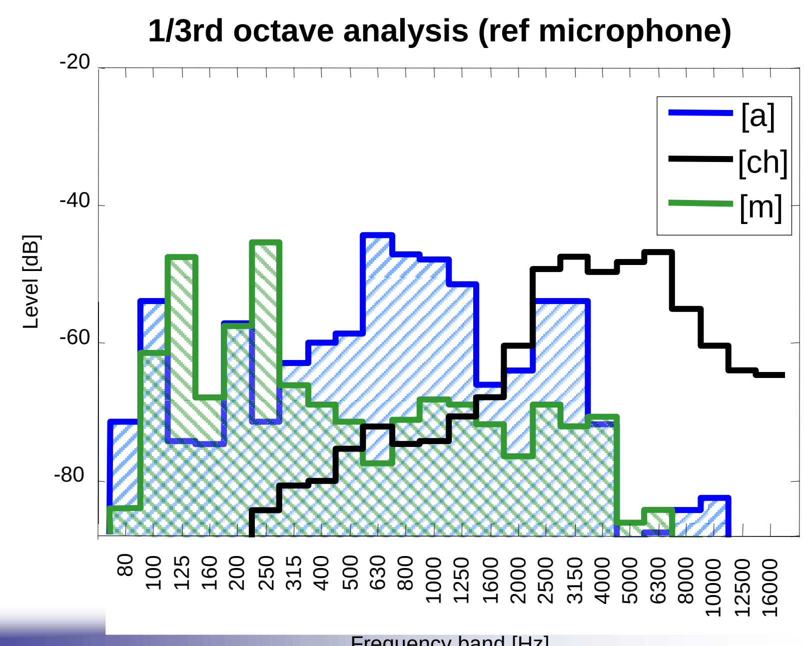


- Differences at mid- and high-frequencies for [f][ch][s]





- Spectral content varies between phoneme



- Due to the large number of frequencies and phonemes to evaluate, it is necessary to develop a more sophisticated analysis technique…

Analysis

- It is apparent that there are clear differences in directivity as a function of frequency and phoneme.
- Due to the large number of comparisons possible, an analysis method is necessary.
- We propose an analysis by comparison of forms followed by a clustering procedure.
  - Comparison of form : cross-correlation coefficient between directivity patterns (R)
  - Dynamic range threshold of 20dB used to reduce effect of frequency bands with little energy
  - Cluster analysis : Group according to cross-correlation coefficient values
  - Correlation Threshold of (1-R) < 0.2 for cluster grouping (difference not significant)

Directivity patterns are **significantly different** if their cross-correlation coefficient, R < 0.8

Cluster analysis (1/3) [1-R]

- Frequency regions where there is no significant difference between phonemes

  80 – 500 Hz

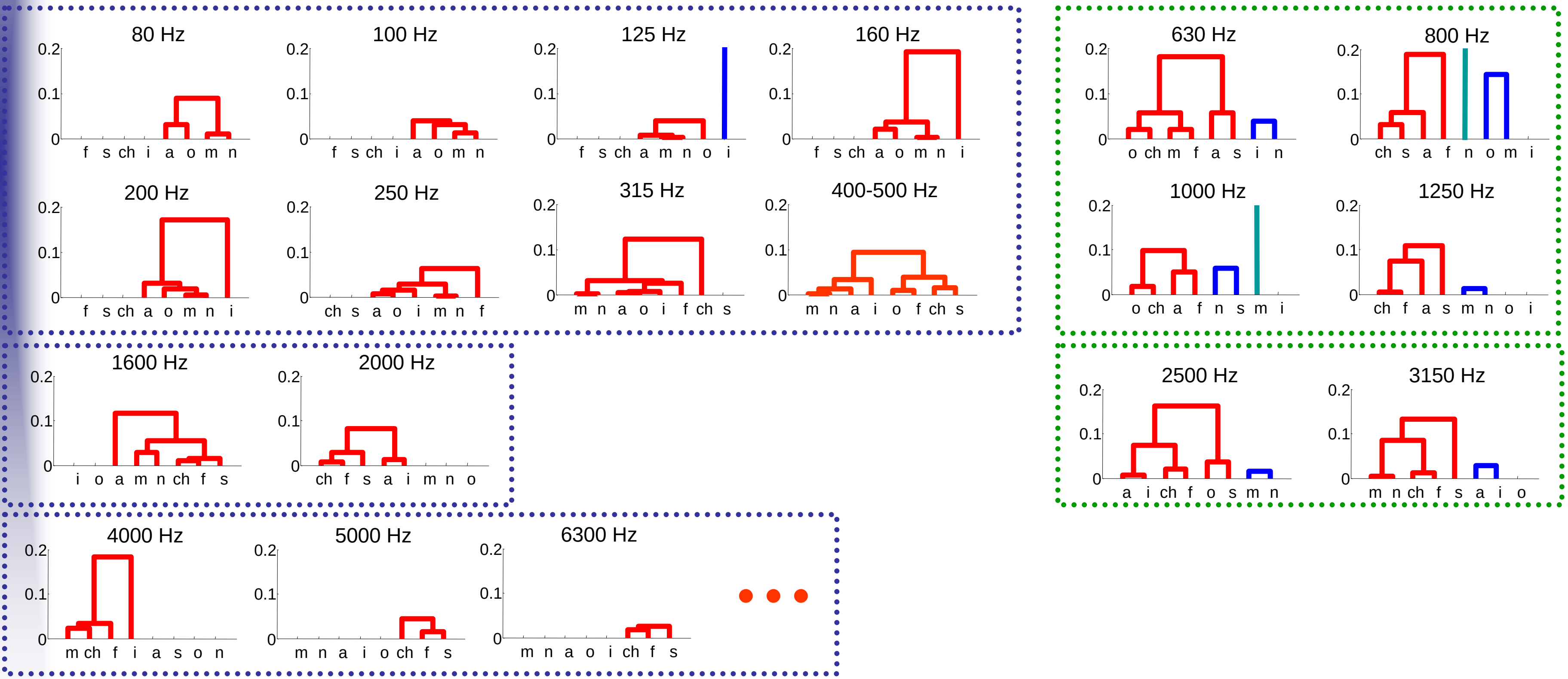
  1600 – 2000 Hz

  4000 Hz +

- Frequency regions where there is significant difference

  630 – 1250 Hz

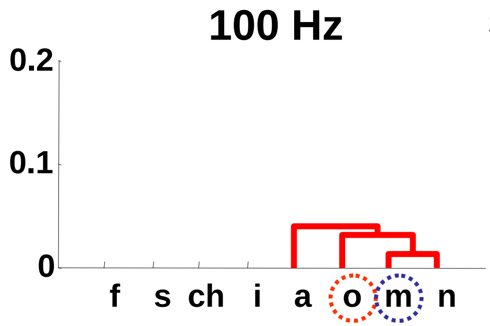
  2500 – 3150 Hz

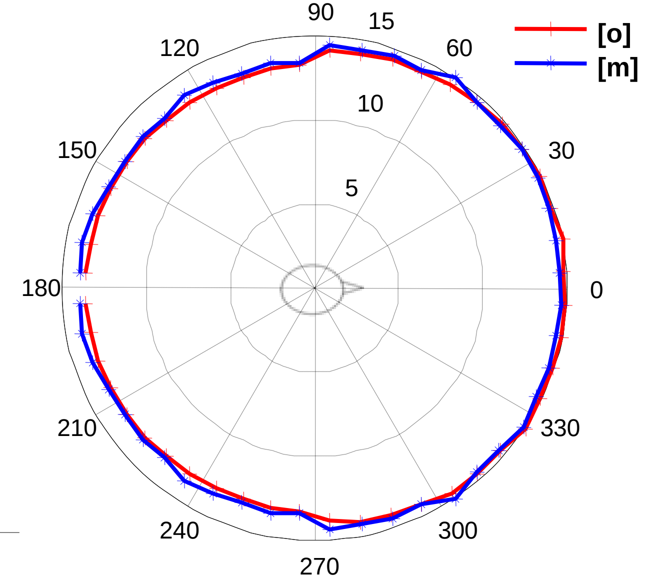


Cluster analysis (2/3)

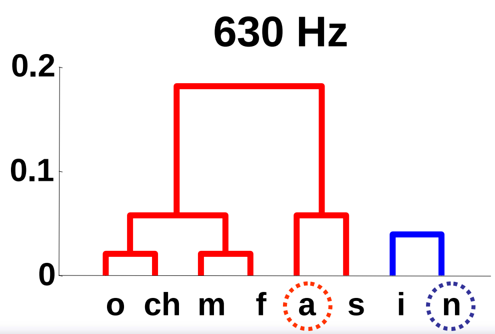
- Representative examples from cluster analyses

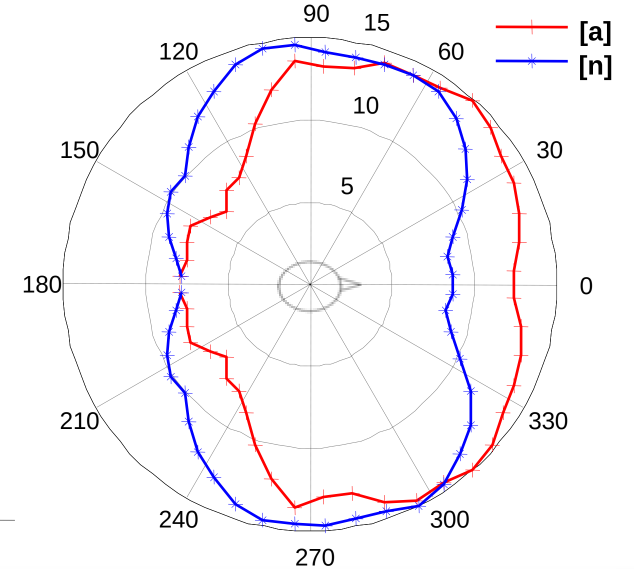
- Low-frequency region where there is no significant difference





- Mid-frequency region where the differences are significant

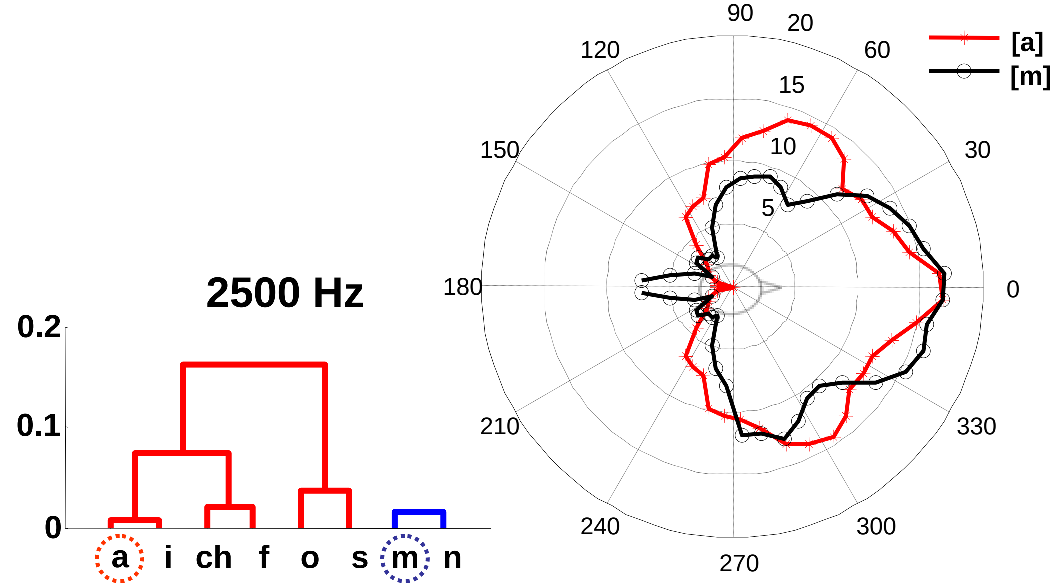




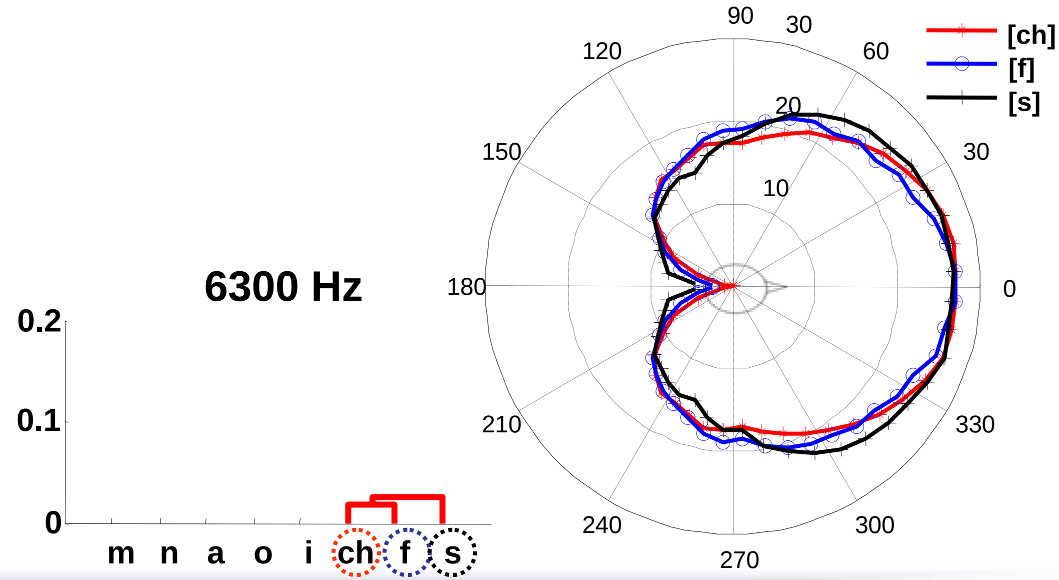
Cluster analysis (3/3)

- Representative examples from cluster analyses (2)

- Mid-frequency region where the differences are significant

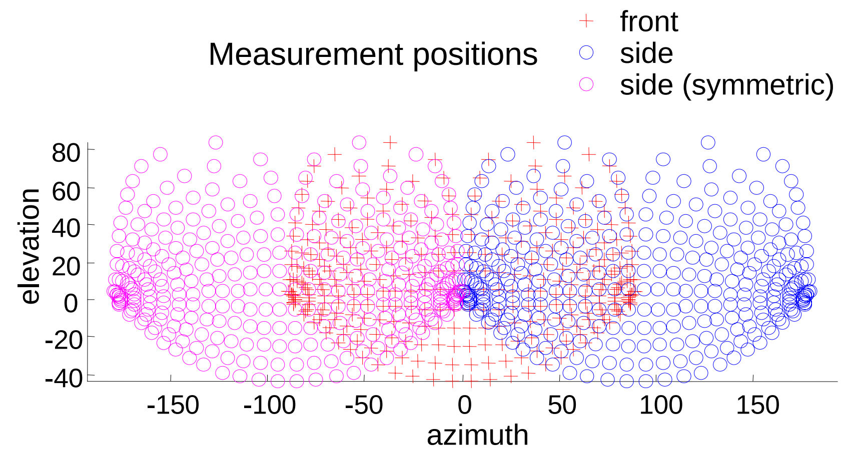


- High-frequency region where there is no significant difference



3D Directivity Preview

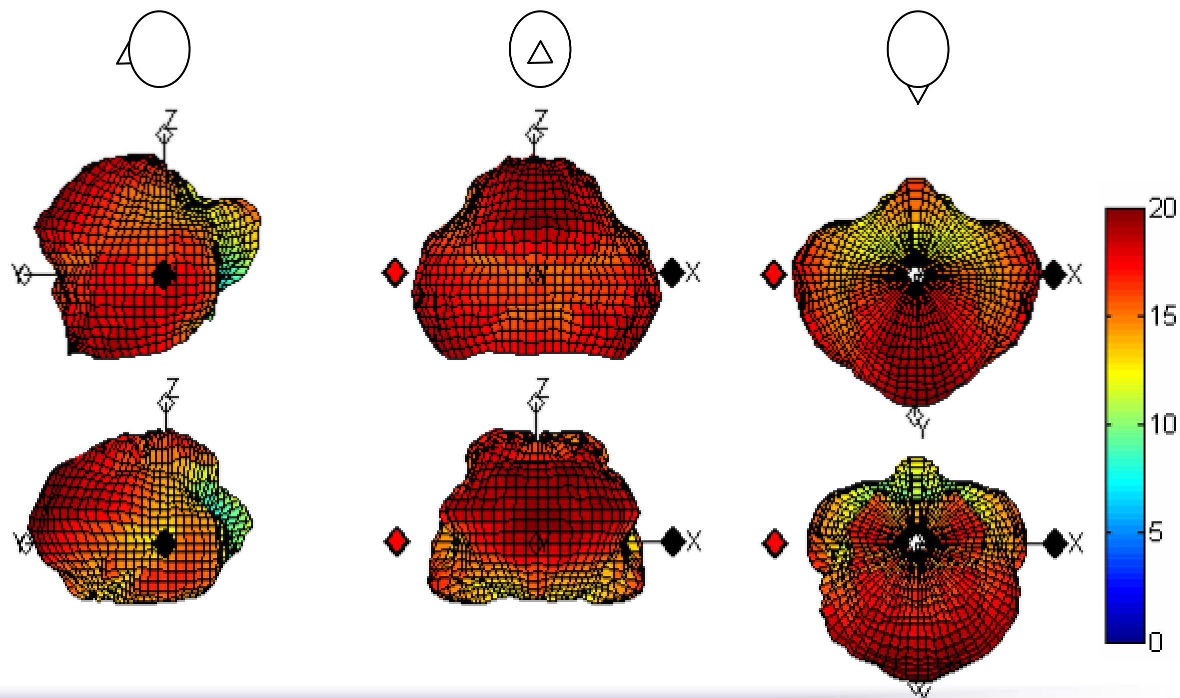
- 3D Directivity patterns are derived by combining the front measured data with the side data, mirrored according to symmetry.



- Resulting data cloud is re-sampled and interpolated on a regular 5° grid.

[a] 1000 Hz

[n] 1000 Hz



Conclusions

- 2D analysis shows significant differences in phoneme directivity patterns
- Cluster analysis identifies frequency ranges where differences are more or less apparent

Future Work

- Investigate correlations between mouth geometry and directivity patterns
- Perform clustering across frequency bands to minimize number of representative patterns
- Additional subjects
- Singing voice directivity : pitch, level, “projection”, and “focusing”

Application Work

- Perceptual study on the detectability / pertinence of results
- Implement directivity patterns into **3D text-to-speech** system for virtual avatars
- Data reduction of directivity patterns via spherical harmonic decomposition