The difference between a question and a statement: A survey of British English Dialects

This poster was presented at the 75th anniversary meeting of the Acoustical Society of America, 24-28 May 2004, New York. It may be referred to as:

Greg Kochanski, Esther Grabe and John Coleman, "The difference between a question and a statement; a survey of British English dialects", J. Acoust. Soc. America 115(5) p. 2398 (May 2004).

http://kochanski.org/gpk/papers/2004/200405ASA

Greg Kochanski

Esther Grabe

John Coleman

University of Oxford, England

GOALS

- Question vs. Statement
  - A basic distinction between two dialog acts
  - Important to the other party in a dialog: they need to know how to respond.
- Therefore one expects an acoustic difference
  - In all dialects.
  - No reason to expect subtlety.
- What is the difference?
  - In F_0?
  - In other acoustic properties?

The Data
(geography)

Belfast

Dublin

Newcastle

Leeds

Bradford
Punjabi

Cambridge

London
Caribbean

The Data (summary)

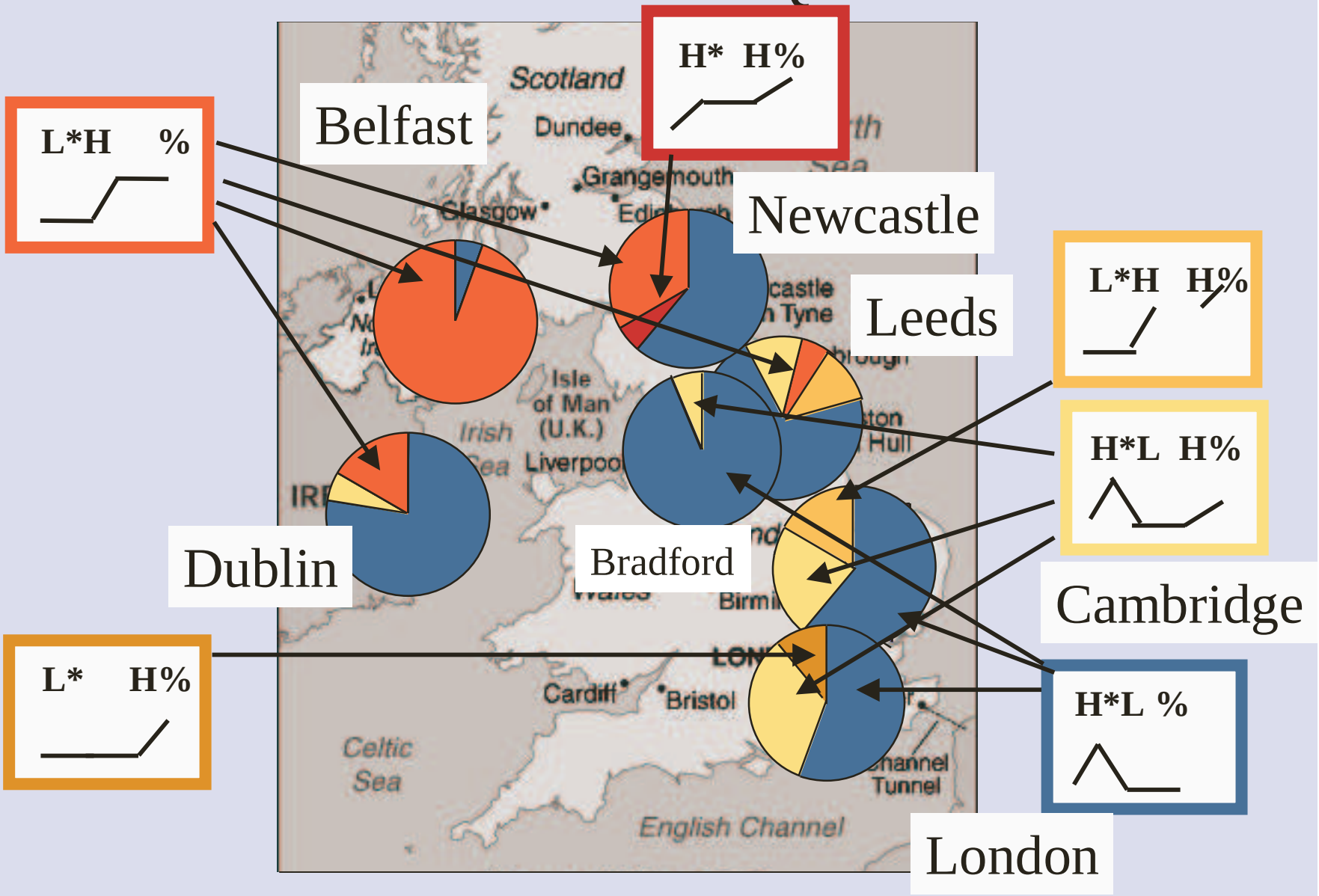
IViE corpus (Grabe, Post, Nolan, Farrar 1997-2002).

- 7 Urban dialects from the British Isles.
- 6 speakers per dialect. Age=16 years.
- Laboratory sentences, read text and retold story, map task, free conversation.
- 36 hours of audio, of which ~3h is labeled in various ways.
- Available on the internet at http://www.phon.ox.ac.uk/~esther/ivyweb/

ESRC (European) Grants R-000-23-7145 , RES-000-23-1049

Background: Final Rises are not inevitable in questions.

Final Accents in Wh-Questions



Background: In all dialects:

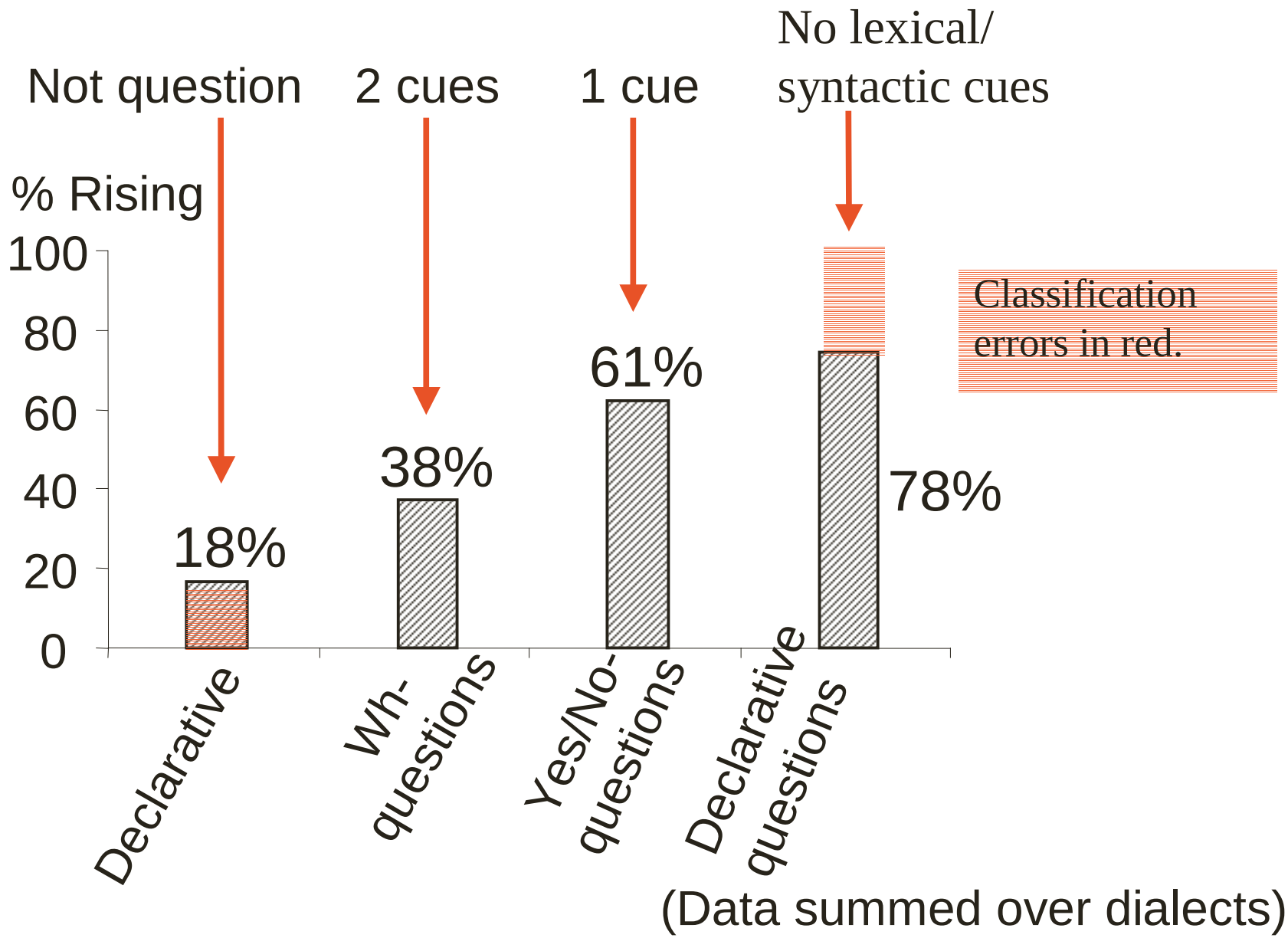
The frequency of final rises was affected by the number of syntactic or lexical cues to interrogativity in questions.

Four types of utterances (read speech):

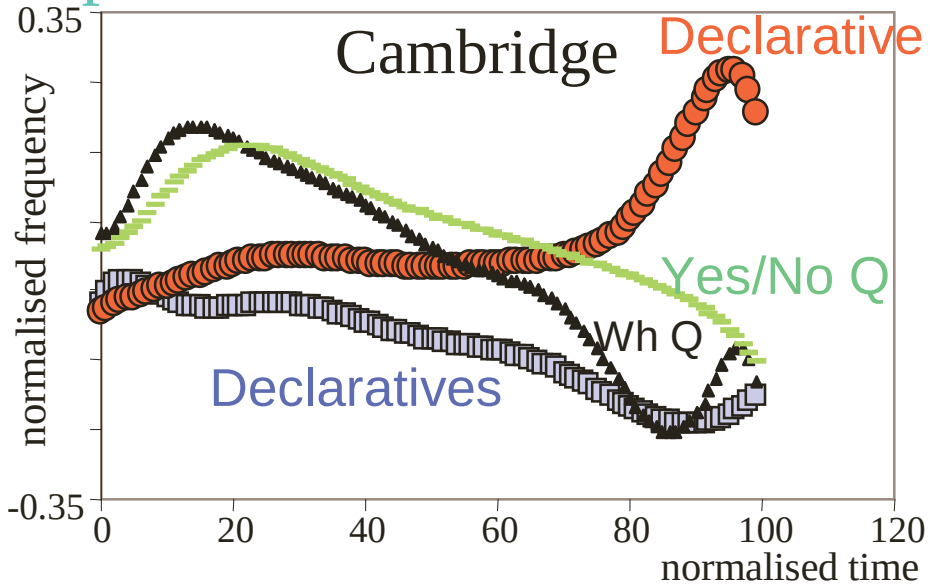
- Declaratives
- Wh-questions [2 cues: word order and “who”, “what”, …]
- Yes/no (polar) questions [1 cue: word order]
- Declarative questions [0 lexical/syntactic cues]

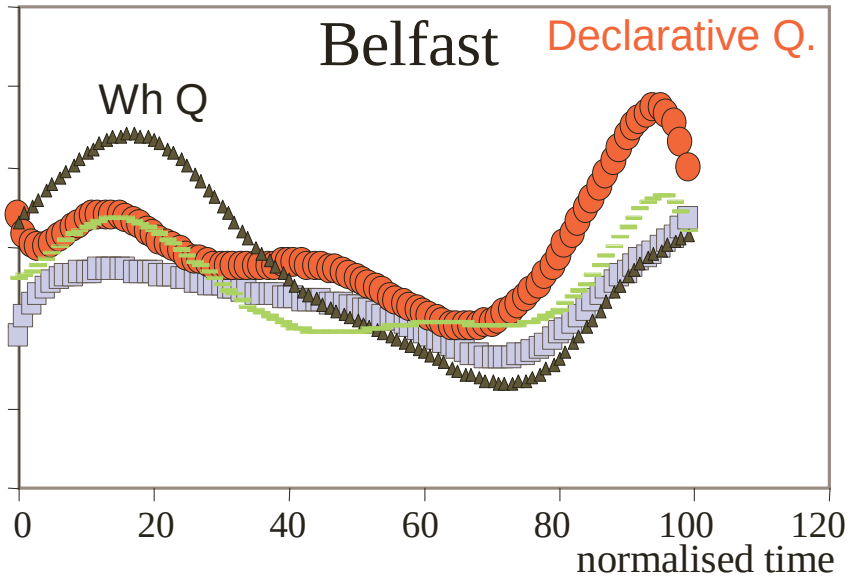
•acoustic cues must be important.

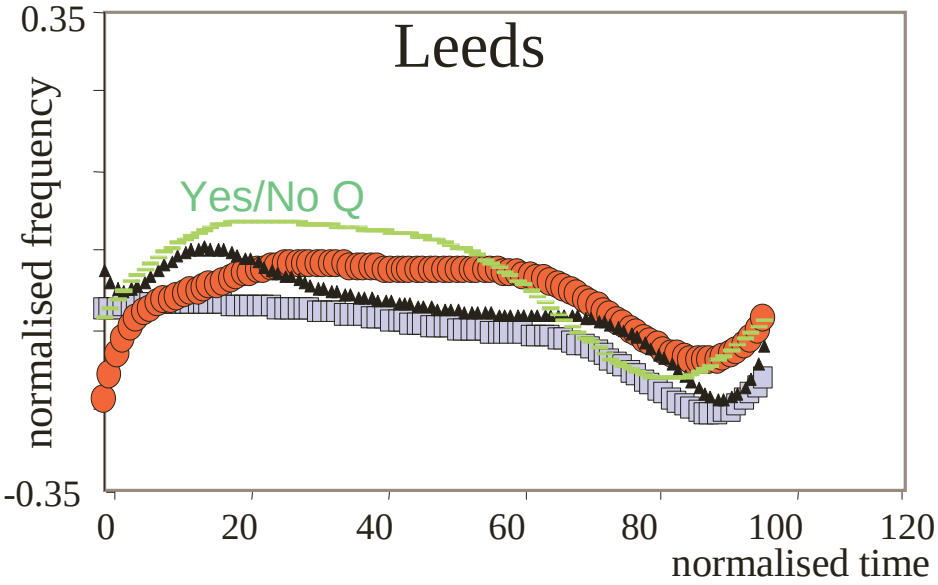
Background: Correlation between cues to interrogativity and final rises

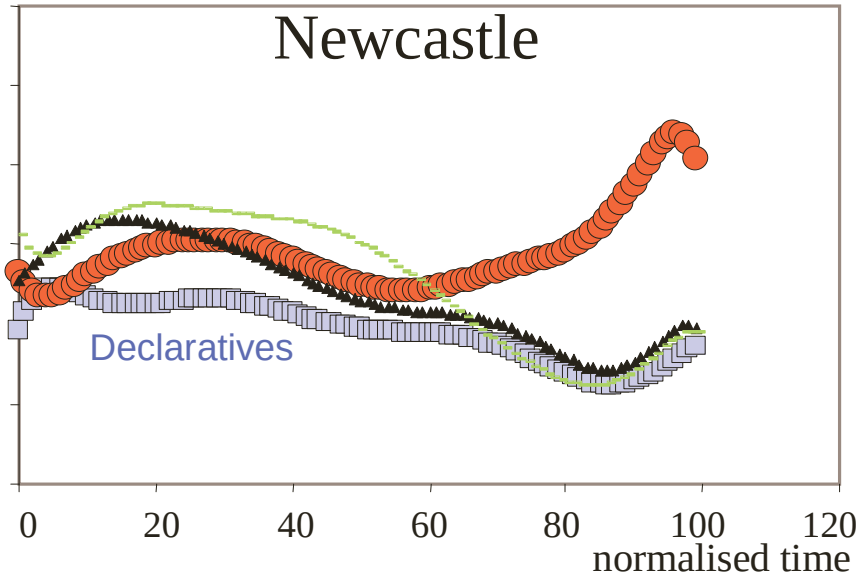


Shapes of different dialects







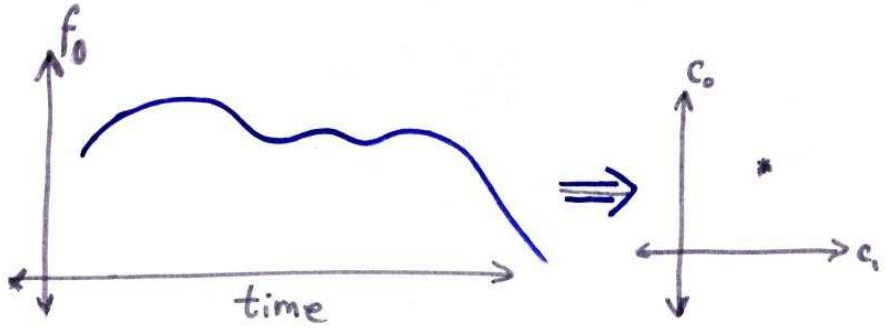


Final rises

- Dialects are very variable in how they use f_0 to mark a question:
  - Some dialects use almost no final rises
  - Some dialects (Belfast) use final rises almost everywhere.
- Even inside a dialect, sometimes final rises are used, and sometimes not.

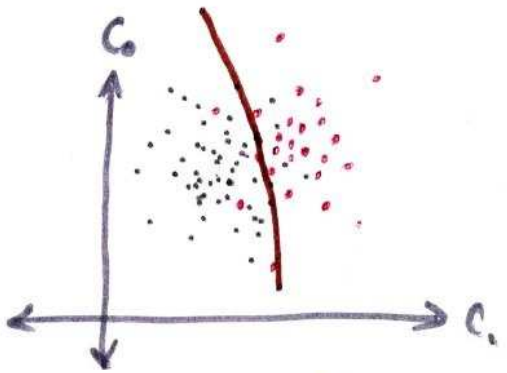
Building classifiers to obtain quantitative results:

1. Map the time-series of (*e.g.*) f_0 into a point (in a multi-dimensional space).

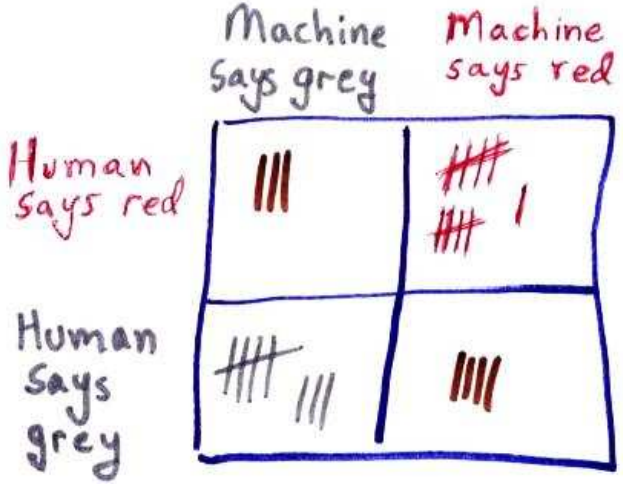


2. Make a cloud of points, one for each utterance or syllable.

3. Find how to divide that cloud to best reproduce the human classification.



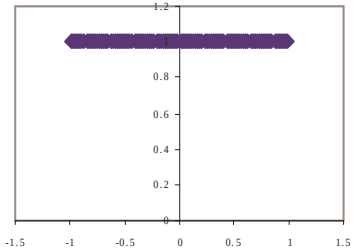
4. The accuracy of the machine classification is a measure of how much information is carried by each acoustic property.

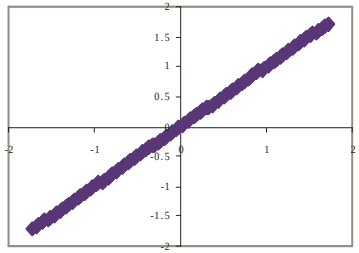


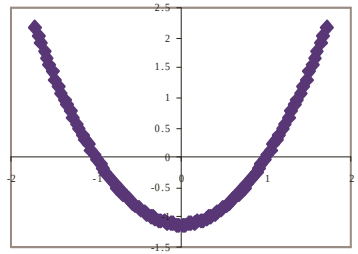
Despite the variation in contour shapes,
is there anything the dialects have in common?

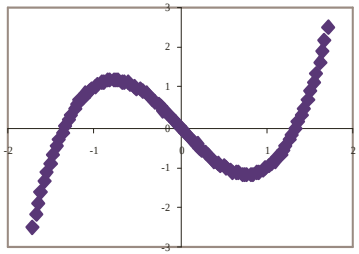
Can we find out using the orthogonal polynomial models?

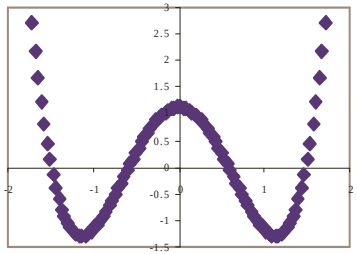
The first Six Legendre Polynomials:

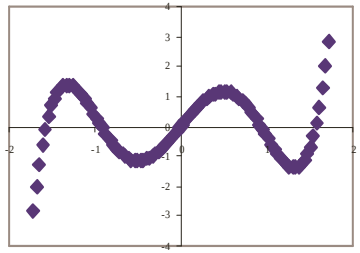
P_0 

P_1 

P_2 

P_3 

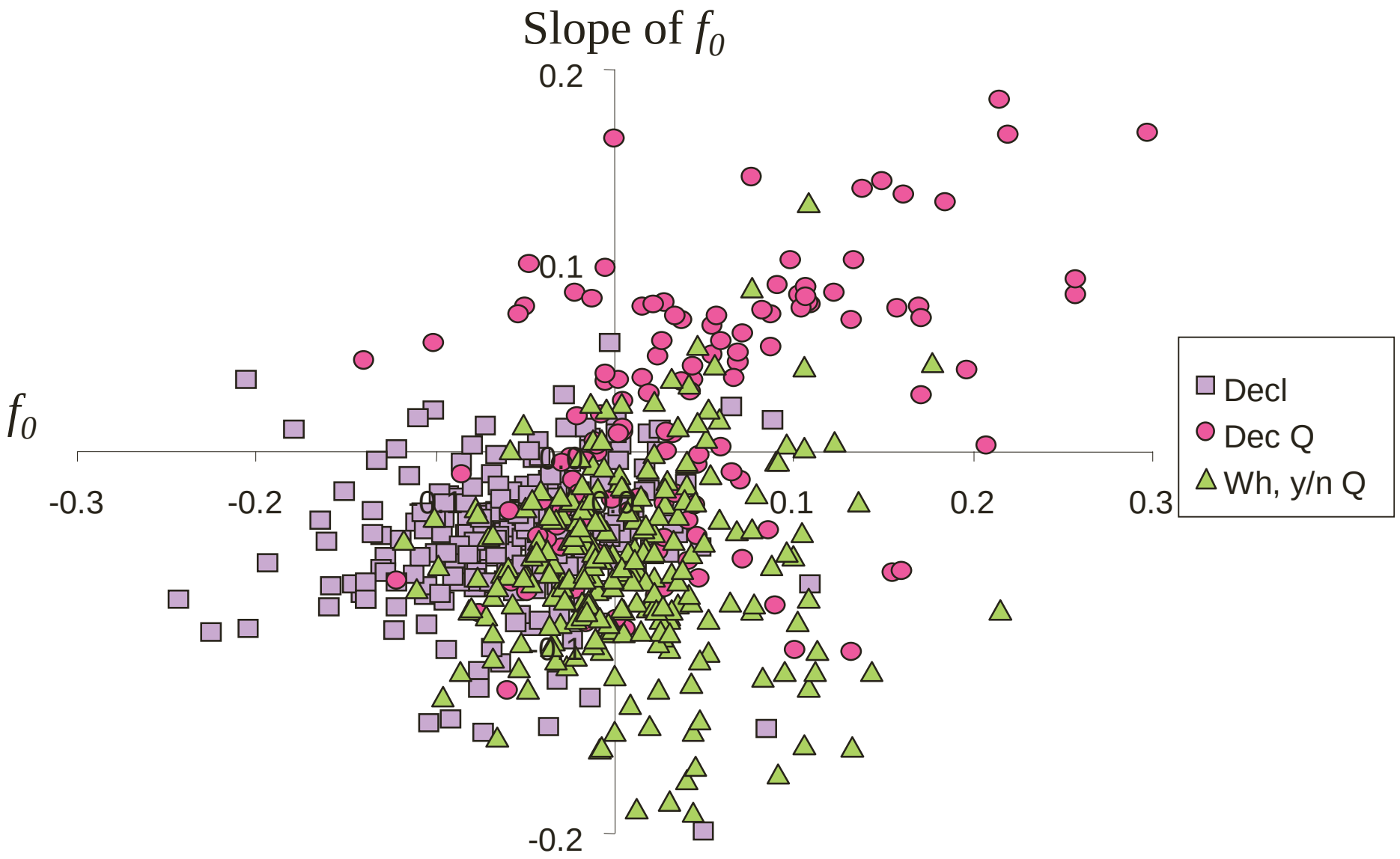
P_4 

P_5 

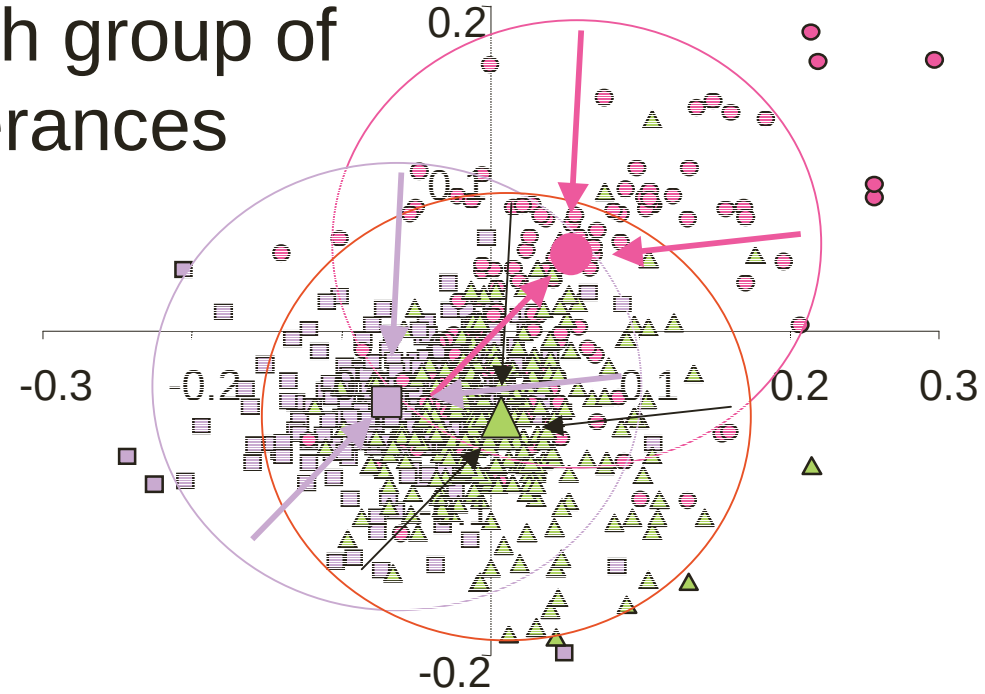
Distinguishing features in f_0:

- As an example, let's look at a maximally simple classifier on f_0 data:
- If we just look at average and slope:
  - how well do average and slope distinguish between the four utterance types?
  - are there dialect differences?

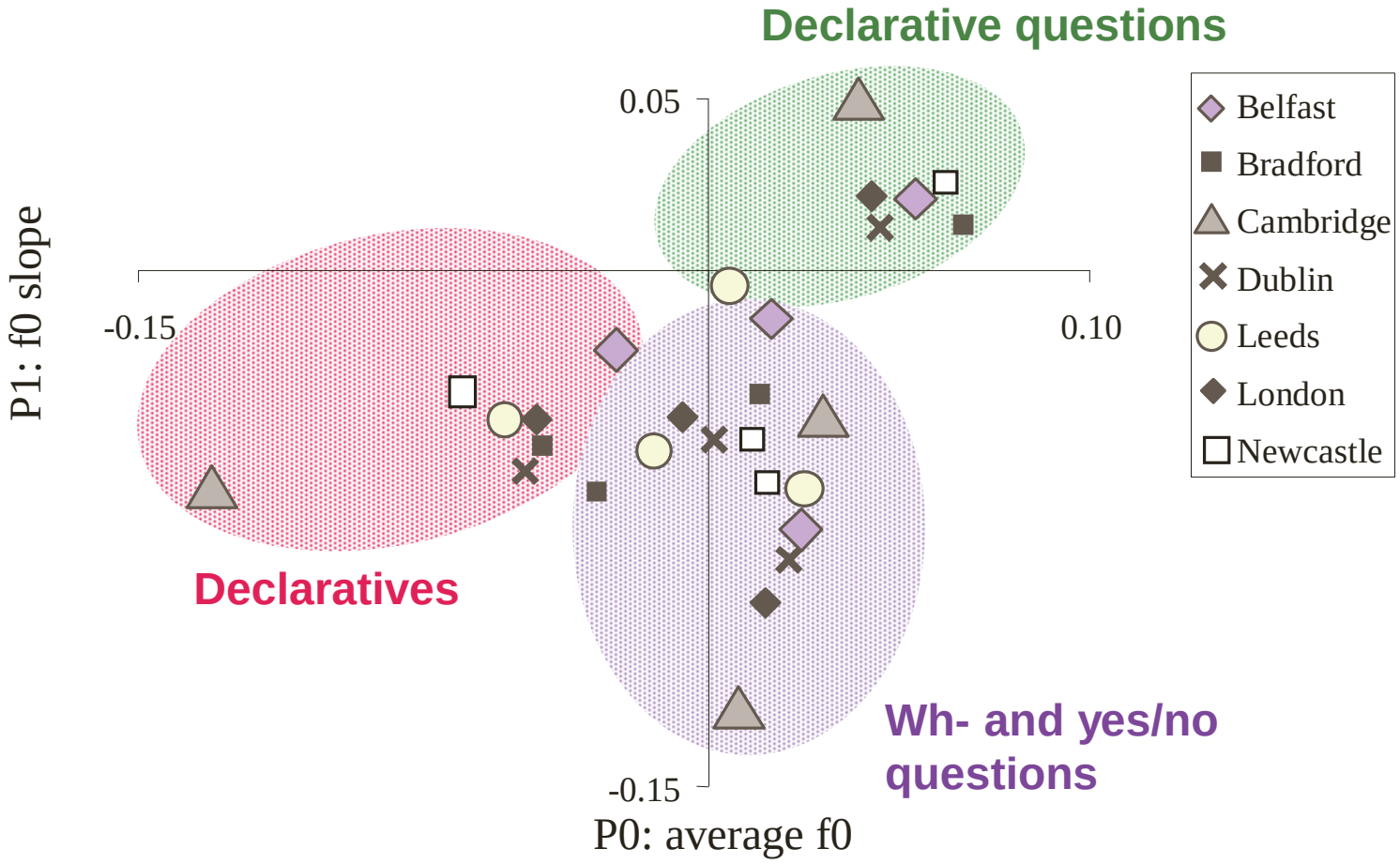
Separation of utterance types by average f_0 and f_0 slope



Compute the median of each group of utterances



Average f_0 and f_0 slope in seven dialects.



Each symbol represents one dialect. Symbols are median points of all utterances in a dialect.

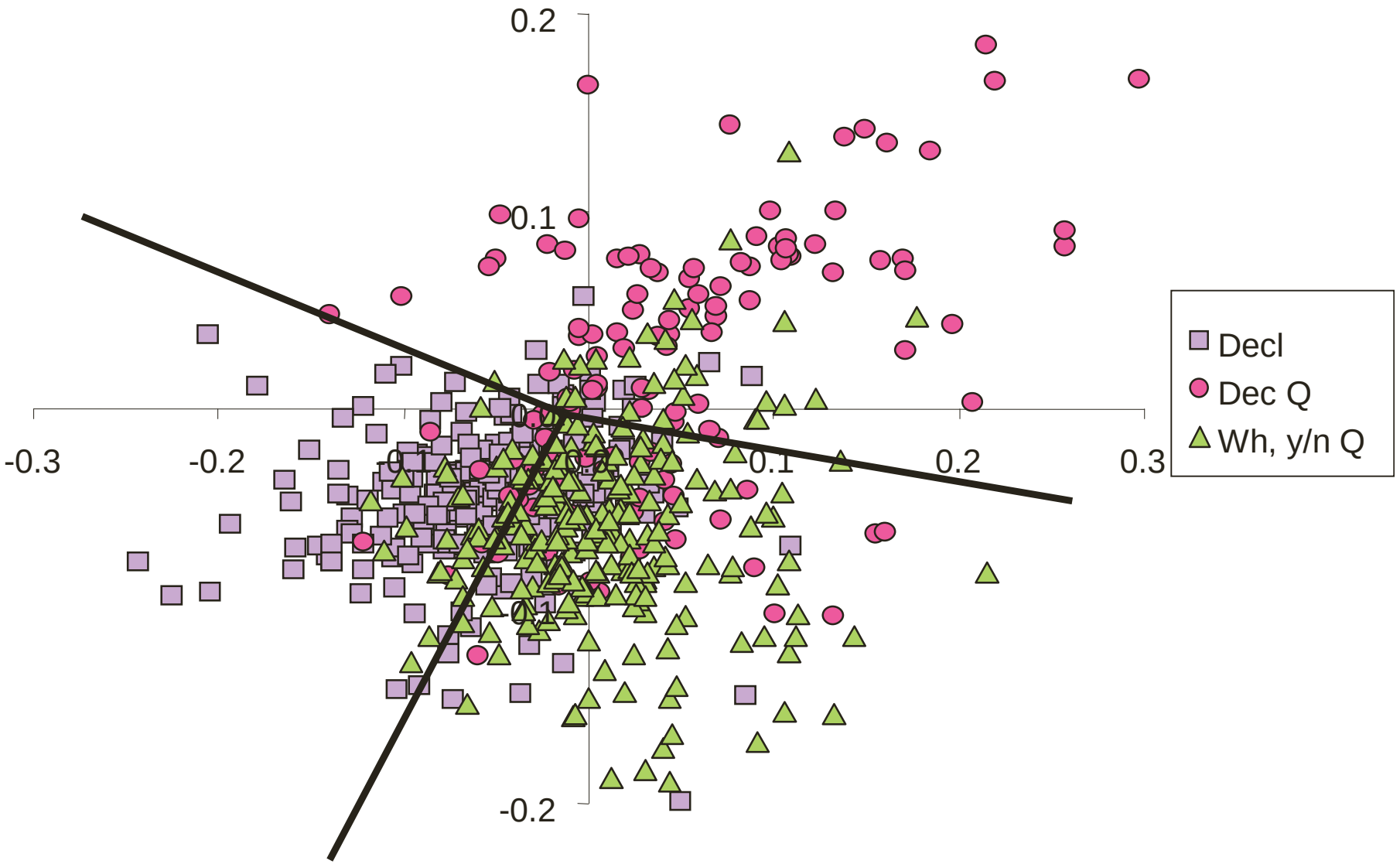
In all dialects:

- average and slope go a long way towards distinguishing between utterance types. Additionally,

- Average f_0 of wh-, yes/no and declarative questions is generally higher than average f_0 of non-questions.

- Average alone goes a long way towards distinguishing questions from non-questions.

Separation of utterance types by average f_0 and f_0 slope



Other Acoustic Features

- How much do other acoustic features contribute to the 2-way or 4-way discrimination?
  - Loudness (ISO R-432, Stevens)
  - Spectral slope (High freq power vs. low freq power)
  - Uniformity (Given a 20 ms long window, how precise a match can one find nearby?)

F_0 and loudness

***F_0*:** Standard measurement, except data weighted towards loud regions of high uniformity (*e.g.* vowel centers).

Loudness: *Not* the same as RMS amplitude.

Uses Steven's Mark VII (1971) algorithm to emulate perceptual judgments. (Mark VII is an improved version of the ISO-R532 Method A standard noise measurement.)

Differences from RMS:

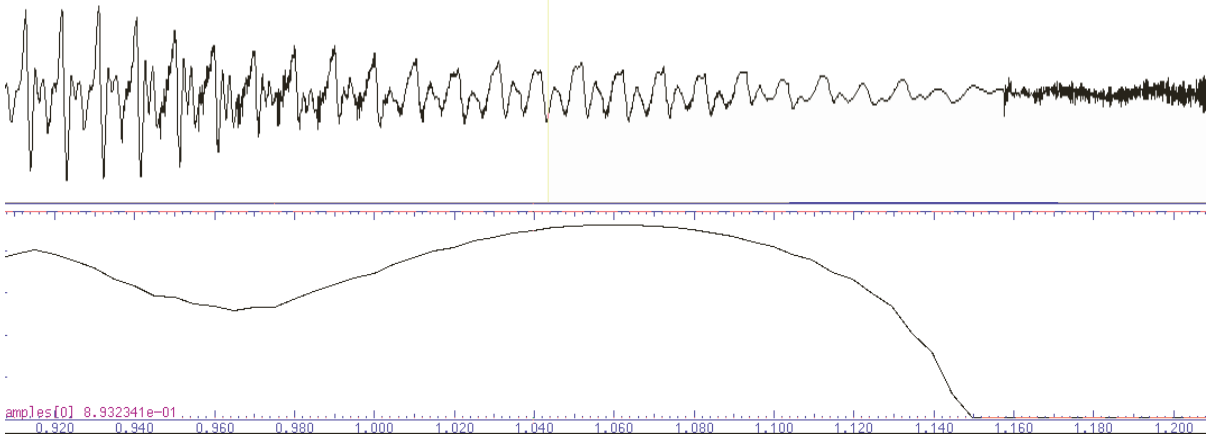
•Close to human perception.

•Takes information from a broad range of frequencies, rather than primarily from the lowest harmonic.

•More stable; not strongly affected by accidental alignments of f_0 and F_1.

•Algorithm loosely analogous to early stages of human auditory system.

Measures of timbre – voicing uniformity

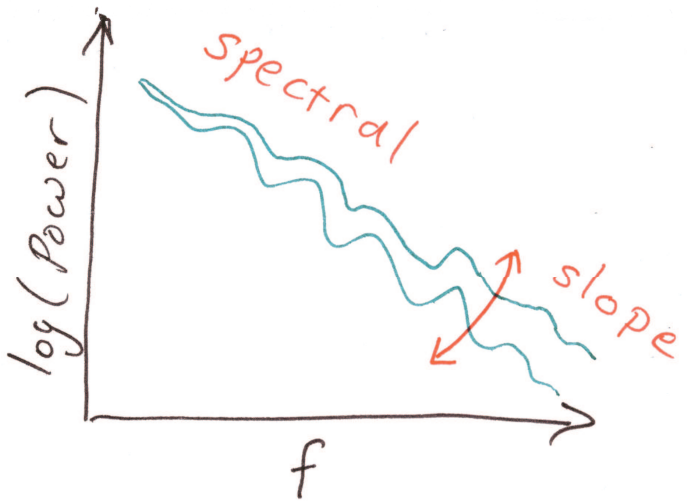


Voicing uniformity: measures the extent to which nearby voicing pulses are identical. Expected sensitivity:

- Glottalized(0) ⇔Modal(1),
- Breathy(small) ⇔ Modal(1)
- Fricative(0) ⇔ Voiced(1)
- Unstable formant structure(-) ⇔ Stable formant structure(1)

Measures of timbre – spectral slope

Spectral slope: measures relative amounts of high (2-5 kHz) and low (0.5-2 kHz) power.

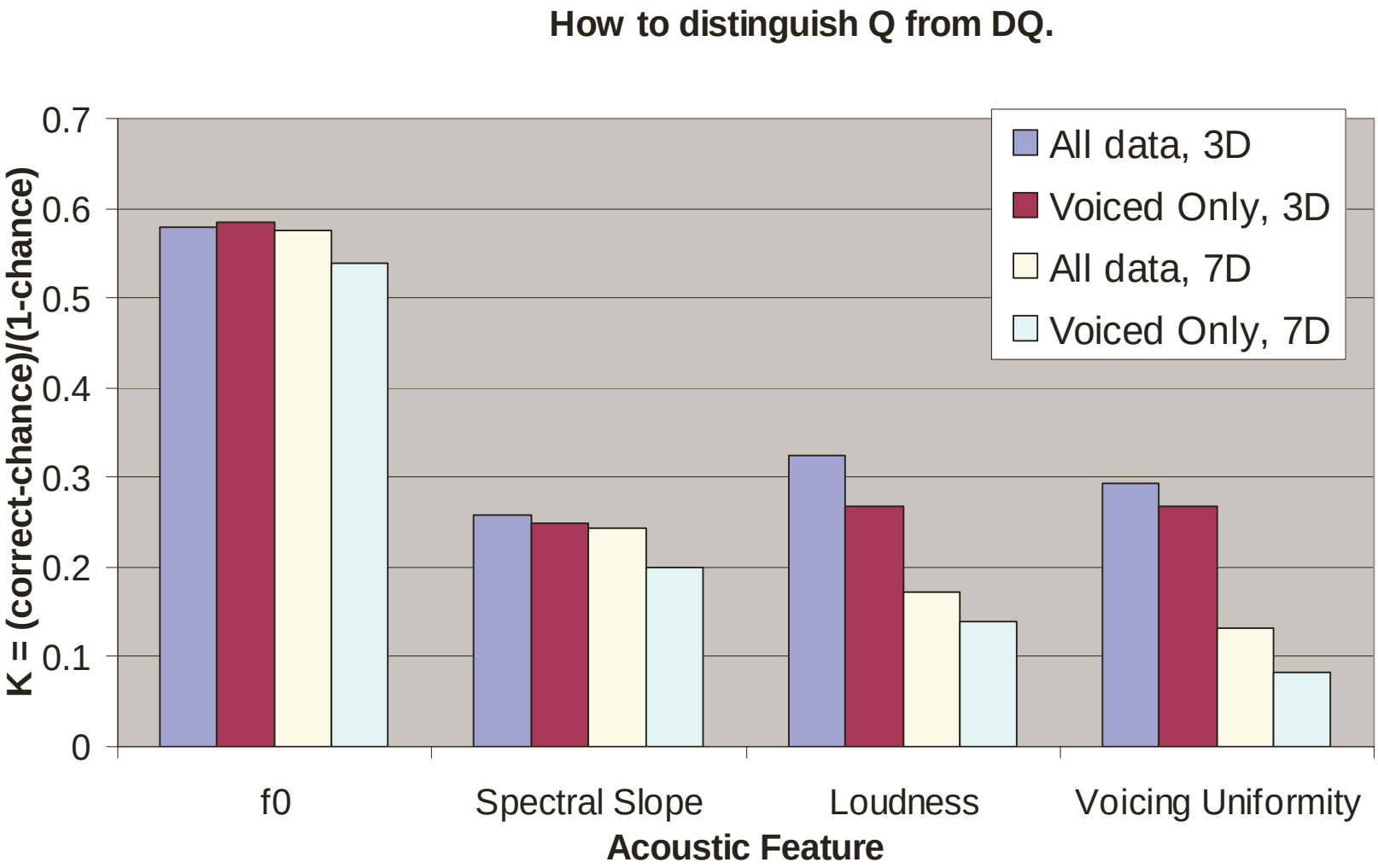


Expected sensitivities:

- Soft(-) ⇔Harsh(+),
- Modal(-) ⇔ Breathy(+)
- Voiced(-) ⇔ Fricative(+)
- Low subglottal pressure(-) ⇔ High subglottal pressure(+)

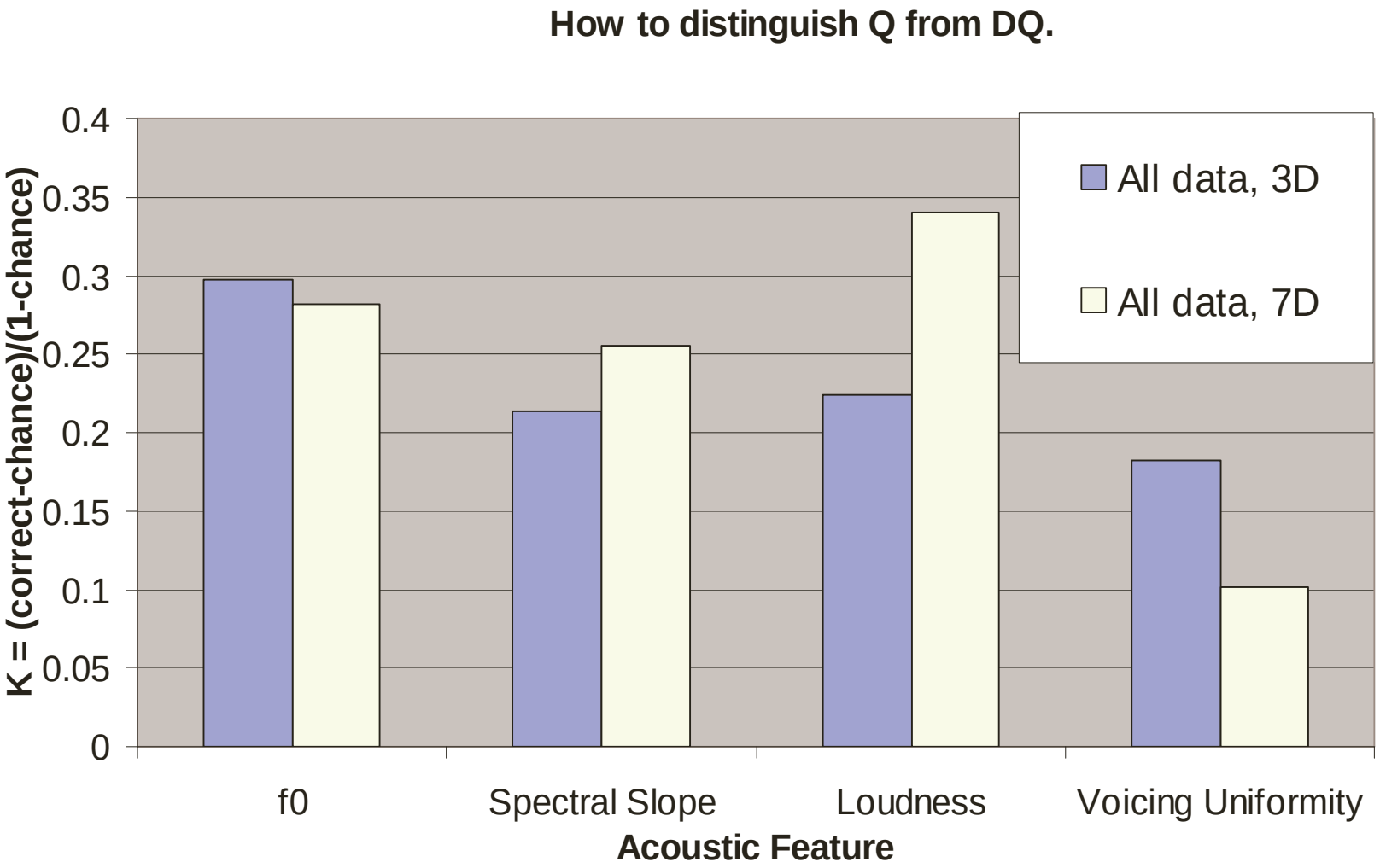
Which acoustic properties matter for the 2-way Declarative vs. Declarative Question distinction?

[K is proportional to how often a particular acoustic property leads to the correct (*i.e.* human) classification.]



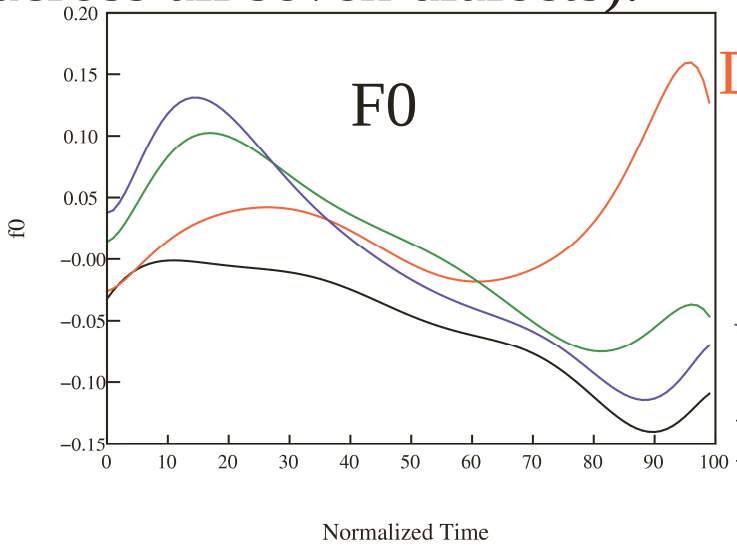
Which acoustic properties matter for the 4-way D *vs.* WHQ *vs.* Yes/No Q *vs.* Declarative Question distinction?

[K is proportional to how often a particular acoustic property leads to the correct (*i.e.* human) classification.]



Median time-series of all four acoustic measures

(across all seven dialects).

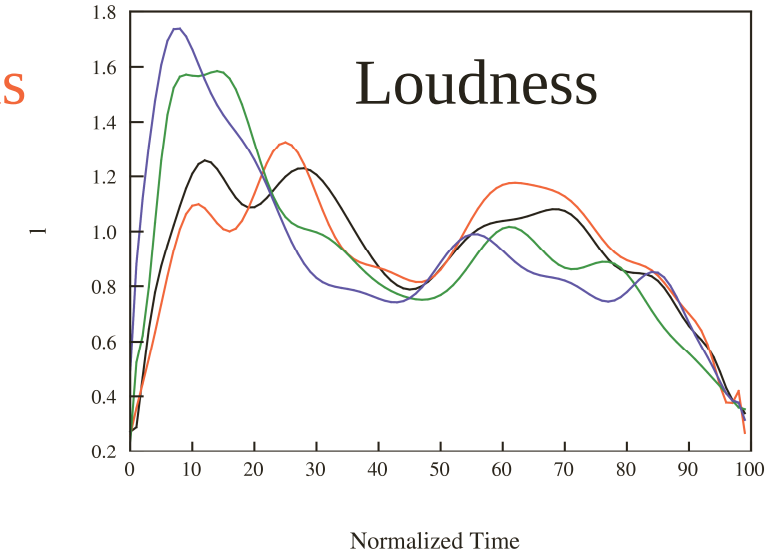


Decl. Questions

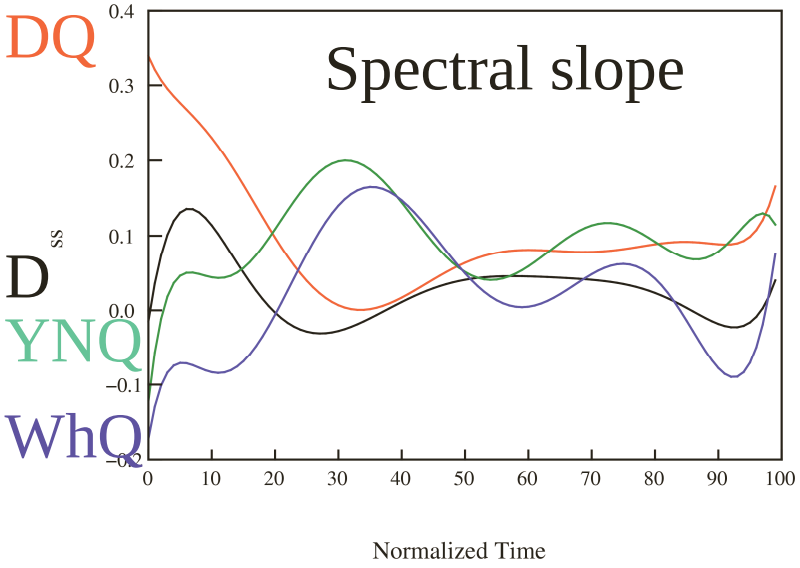
Yes/No Q.

Wh Questions

Declaratives



DQ



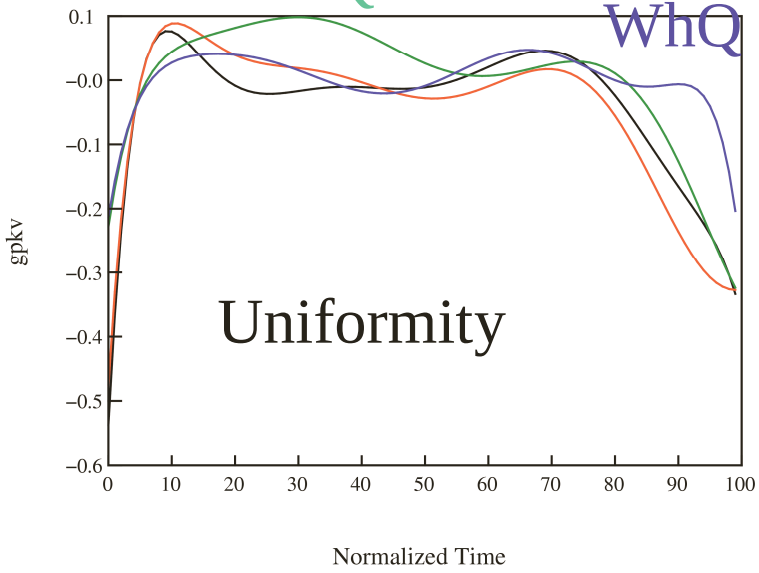
D

YNQ

WhQ

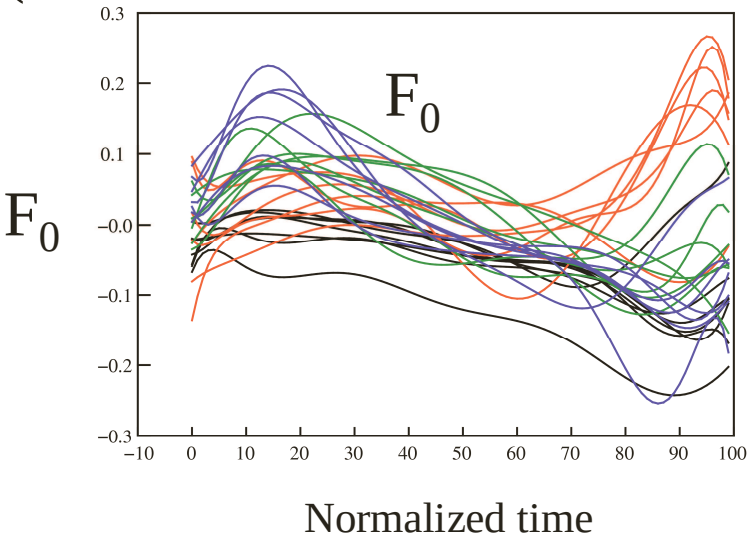
YNQ

WhQ



Median time-series of all four acoustic measures

(each curve shows median of one dialect).

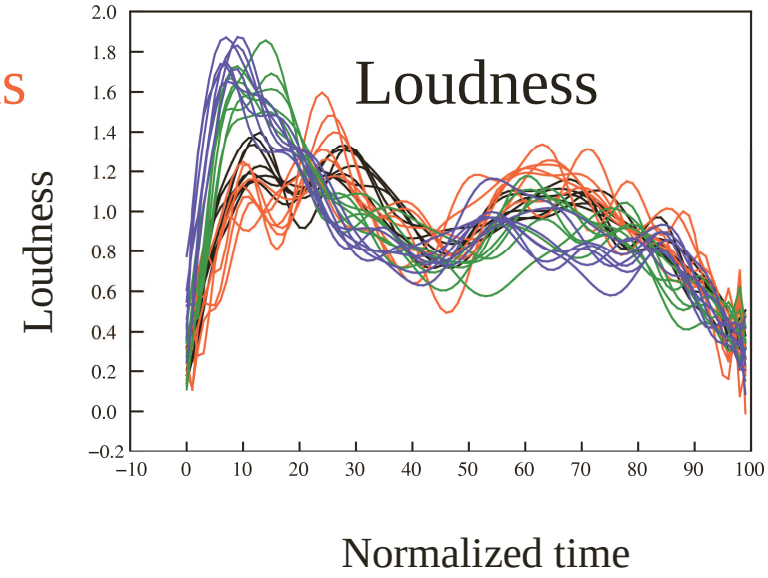


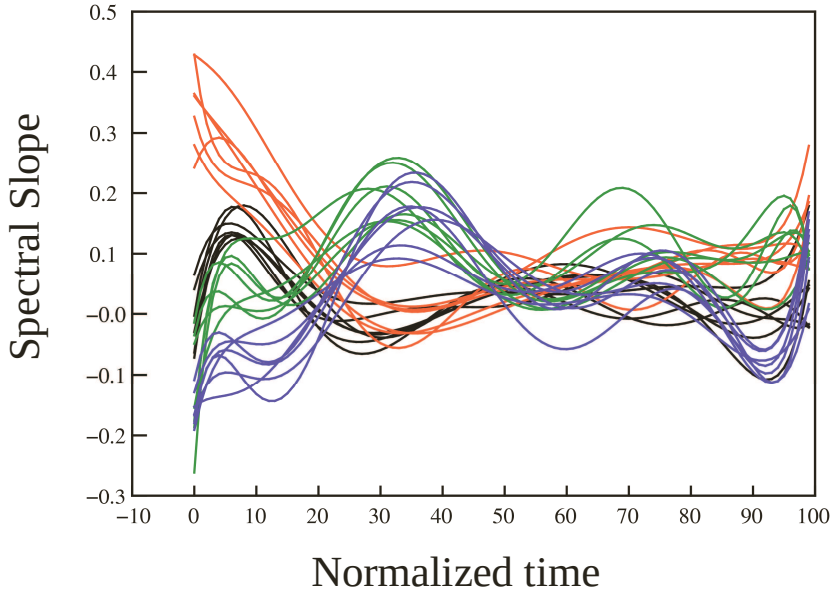
Decl. Questions

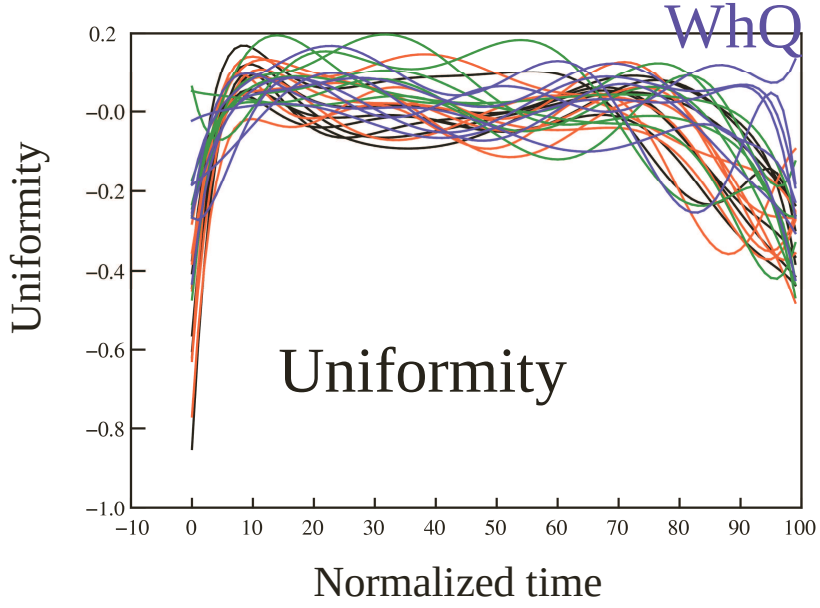
Yes/No Q.

Wh Questions

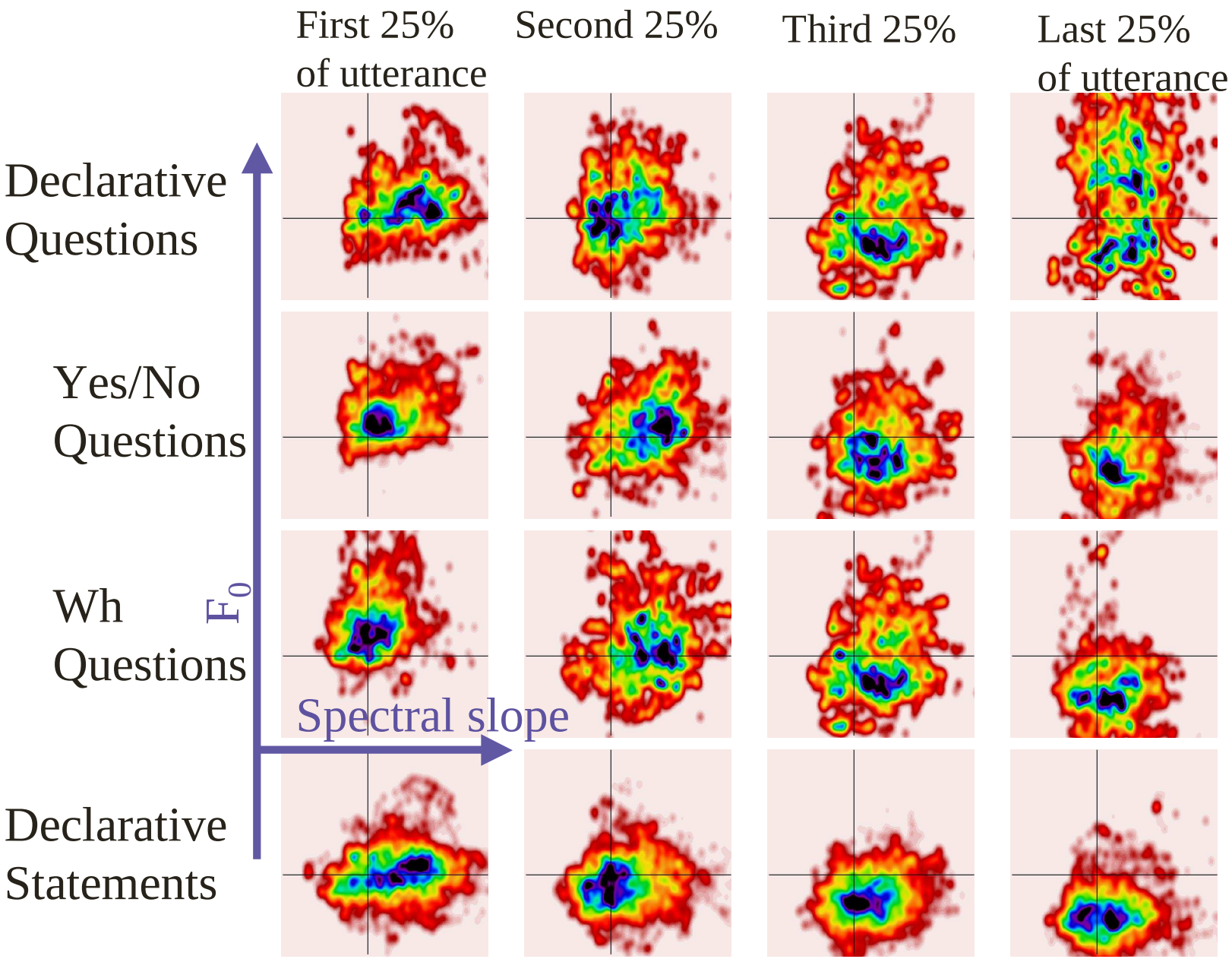
Declaratives







F0 and Spectral slope (2-D histograms) vs. time.



Conclusions

- Dialects can have very different tune shapes, but they make similar question/statement distinctions.

- Old reliable f_0 …
  - … is indeed the best way to make the 2-way D/DQ decision.
  - …doesn't tell you much about the 4-way distinction.
- Other acoustic properties…
  - …carry a substantial amount of information.
- The 2-way Declarative/DQ distinction is…
  - … clearest at the end, but
  - …spread across the utterance, including the very beginning (*e.g.* spectral slope)