Getting a handle on category membership in intonation
Intonation

• Intonation plays central role in human communication
  – it can provide immediate cues to start new word or phrase in the speech stream, and to meaning of utterances
  – research in L2 learning, aphasia, speech synthesis and automatic speech recognition shows that when intonation is wrong, communication often breaks down

• But there is no agreement about
  – how intonation is realised in speech
  – what linguistic units are involved
  – what role it plays in speech comprehension, or
  – how it is represented and processed in the brain.

• Why? Intonation is continuous in nature: unclear how acoustic signal maps onto meaning
This talk

- Understanding the role of intonation in human communication demands a multidisciplinary approach which crosses traditional academic boundaries

- ESRC-funded project (RES-061-25-0347)
  
  *Categories and gradience in intonation: Evidence from linguistics and neurobiology*

- Project team:
  - Brechtje Post (PI, RCEAL)
  - Emmanuel Stamatakis (CI, Anaesthesia)
  - Francis Nolan (CI, Linguistics)
  - Toby Hudson (RA, RCEAL)
  - Iwo Bohr (RA, RCEAL)
Main objectives of the project

1. to use a combination of empirical paradigms to test the central tenet of the Autosegmental-Metrical framework for intonation analysis (currently the predominant theory)

2. to pin down the neural architecture that supports the processing of the intonational information at issue in a series of speech production and perception experiments and 2 fMRI experiments
Overview of the talk

• What is intonation?

• Autosegmental-Metrical tenet: Phonetics distinct from phonology in intonation

• The problem: How to decide about category membership in intonation?

• The answer: A multidisciplinary approach, combining evidence from linguistics and neurobiology to differentiate phonetic and phonological information in intonation
  – Processing hierarchies in speech perception
  – fMRI experiments
  – Production and perception experiments

• Expected outcomes
What is intonation?

- Variations in pitch
- Intonation contour: series of changes in pitch

Anna.  Anna?  Anna!?

3 pitch movements which signal different meanings
Functions of intonation

Signal linguistic function utterance:

• Grammatical function of an utterance
  • C'est bien ce que vous dites
    ‘What you say is right’
    ‘It is indeed what you say’
  • Grammatical function of a word (e.g. convict)
  • Speaker attitude / Emotion (paralinguistic)
  • Information structure (linguistic)
  • Focus (linguistic)
So how do we go from sound to meaning in intonation?

- **Categorical interpretation**: At some stage in the comprehension process, some of the continuous information is interpreted categorically and decoded into distinct meaningful units, e.g. a rising pattern that marks a question.

- **Gradient interpretation**: Sometimes intonation makes a more gradient contribution to meaning, e.g. when gradual increases or decreases in a particular feature like pitch convey a more angry or less timid tone of voice.
Problems

• Functional categories and gradient variation are closely intertwined in intonation, since both can be used to signal linguistic as well as paralinguistic variation in meaning (e.g. Crystal 1969, Bolinger 1970).
  – Categorical variation in form can signal paralinguistic meaning (e.g. rise vs rise-fall for surprise)
  – Gradient variation in form can be categorically meaningful (e.g. height of a rise contrasting questions vs. continuations)

• There is no minimal pair test to decide category membership (as for segments, e.g. tear vs beer)

• We know very little about how acoustic cues interact in signalling different ‘types’ of meaning
Intonational realisation: Multiple cues

- Pitch
- Duration
- Pauses
- Relative loudness
- Spectral properties
- Voice quality

E.g. Increases in energy accompanied by
- higher pitch
- greater loudness
- longer duration
- changes in spectral slope
- stronger obstruent releases

(Vaissière 2005)

Co-occur in the signal to

Enhance each other or interact with each other in signalling intonational meaning (see Post et al 2007 for overview)
K: *pas euh vraiment droit, mais un petit peu euh [...]*
   Not really straight but a bit eh [...] 
I: *Oui*
   Yes 
K: *un petit peu en arrondi sur la gauche,*
   A bit round to the left 
I: *En arrondi qui tourne à gauche?*
   Around turning to the left? 
K: *Non non, tu tournes pas à gauche, tu continues toujours comme si t’allais tout droit, mais ça fait un petit virage, quoi.*
   No no, you don’t turn to the left, you carry on as if you’re going straight, but it makes a little bend, [sort of.] 
I: *D’accord.*
   All right.
Multiple interacting cues: Turn-taking in French

Non non, tu tournes pas à gauche, tu continues toujours comme si t’allais tout d

pas euh vraiment droit, mais un petit peu euh

3 syll./sec.

5 syll./sec.

Change in pitch register, speaking rate and loudness
Findings of multiple cues in signalling different speech acts:

• Speaking rate was faster in the questions than the statements in Neapolitan Italian (D’Imperio 2000)

• Pausing strengthens the interrogative interpretation of Swedish utterances (House 2003)

• Voice quality can be more tense on prominent words in interrogatives than in declaratives in English (Epstein 2003)

• Voice quality can be more lax at the end of questions in languages spoken in Africa (Rialland, to appear)

• Breathy voice interacts with speech act in Japanese (Campbell and Moktari 2003)
Summary so far

• Intonation can signal linguistic as well as non-linguistic (‘paralinguistic’) meaning
• Primarily cued by variation in fundamental frequency (but some acoustic variation is extra-linguistic, e.g. microprosodic effects)
• F0 variation co-occurs with variation in other cues, with which it may interact

• Question: how can we model this?
Phonetics vs phonology

• We don’t perceive all differences between sounds
  e.g. [ιæ] in tear ≠ [ιæ] beer

• When we do perceive a difference, we don’t necessarily use it to make a meaningful distinction

  Non-contrastive, gradient phonetic variation

• But some differences are meaningful

  Contrastive, categorical, phonological variation
Modelling intonation: Intonational phonology

• Intonation refers to
  – the use of suprasegmental features
  – to convey sentence-level pragmatic meanings
  – in a linguistically structured way

  (Ladd 1996)

• In the analysis, distinguish between
  – phonological structure and phonetic implementation
  – identification of distinctive patterns and the phonological organisation of their component units (primitives)
Phonetics versus phonology in intonation

- Linguistic function: Choice of intonation contour
- Non-linguistic, meaningful variation (e.g. Attitude, emotion)
- Physiological variation between speakers (e.g. register men vs women)
- Speaking environment (e.g. higher pitch peaks in noise)
- Microprosodic variation (e.g. /i/ intrinsically higher than /a/)

Phonological choice vs Phonetic realisation
### Categorical and gradient variation in form and meaning

<table>
<thead>
<tr>
<th>Factors</th>
<th>Intonation contours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>description</td>
</tr>
<tr>
<td><strong>Phonology</strong></td>
<td></td>
</tr>
<tr>
<td>categorical change in form</td>
<td>linguistic: question rise vs. declarative fall (categorical change in pitch direction)</td>
</tr>
<tr>
<td>categorical change in meaning</td>
<td></td>
</tr>
<tr>
<td>change is meaningful</td>
<td></td>
</tr>
<tr>
<td>stylised F0 contour</td>
<td></td>
</tr>
<tr>
<td><strong>Phonetics</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>linguistic: question vs. continuation rise (higher peaks in questions than continuations)</td>
</tr>
<tr>
<td></td>
<td>paralinguistic: question rise expressing surprise (peak height varies as a function of surprise)</td>
</tr>
<tr>
<td></td>
<td>extralinguistic: question rise against background noise (higher peaks in more noise)</td>
</tr>
</tbody>
</table>
Modelling intonation: Turning points

• Contours are analysed in terms of high and low turning points which align with specific locations in the segmental string:
  - stressed syllables
  - prosodic boundaries

  *Was it *yesterday* that I was looking for an example?*

• Turning points can be represented as tones: High or Low (H or L): two distinct levels

• Whose F0 values are specified in phonetic implementation, depending on speaker characteristics and context

(Pierrehumbert 1980, Bruce 1977, Gussenhoven 1984)
Modelling intonation: Turning points

High or Low turning points

The tonal structure

Only some configurations of tones are contrastive: pitch accents and boundary tones
Autosegmental-Metrical framework: Advantages

- Offers discrete, economical and insightful formalisations of intonation systems

- allowing clear hypotheses about how acoustic variation maps onto meaning, distinguishing between phonological categories and phonetic realisation

- This could provide the key to understanding cross-linguistic, dialectal and stylistic intonational variation

- which in turn will open up new avenues for researching cognitive and neural aspects of intonation processing
Research questions

1. Can we provide independent behavioural and neurobiological evidence to back up linguistic hypotheses about phonetics and phonology in intonation?

2. In doing so, can we provide a more refined, linguistically informed model of the neurobiological underpinnings of intonation?
The auditory system

(http://www.brainconnection.com/topics/?main=anat/auditory-anat2)
Auditory cortex

• primary auditory cortex
• secondary auditory cortex
• posterior auditory field
• anterior auditory field.

Connected with, e.g.:
• Wernicke's area: interpretation of language.
• Broca's area: production of language.

(Left hemisphere)
Geschwind:

- inferior parietal lobule at junction of auditory, visual & somatosensory cortexes
- connected to Broca & Wernicke’s areas
- simultaneous multimodal processing to help classify and label incoming information
Dual-stream model of functional anatomy of language

- Two distinct, functionally specified parallel processing streams:
  - Dorsal stream
  - Ventral stream
- Speech processing: multiple streams involving anatomically separable areas which are connected through multiple pathways
- Support several distinct levels of processing serially and in parallel (e.g. Hall, Hart and Johnsrude 2002)

(Hickok and Poeppel 2007)
Processing hierarchies in speech perception

• Interactions with the dorsal-stream network ensure that speech is perceived categorically (Davis and Johnsrude 2007):
  – successive stages of processing achieve greater abstraction from the acoustic input
  – while maintaining multiple possible interpretations of the incoming signal
  – with higher-order frontal regions modulating activity in lower-order temporal regions

• Differential processing, involving structures in STG and LIFG, shown for
  – segmental contrastive information (e.g. Burton et al 2000, Obleser et al 2004, Eulitz and Lahiri 2005)
  – phonological and morpho-phonological processing (Tyler et al 2005)
  – lexical tone (Gandour et al 2003a,b,c, 2004)
Neural correlates of phonological processing

- Studies of segmental contrast, morpho-phonological info & lexical tone all showed more activation for differences that are phonological/morpho-phonologically meaningful in frontal areas
- No difference in activation when it is acoustic (but STG)

Speech input that functions contrastively is perceived differently, also at neural level

Phonological distinctions have distinct neural correlate
• Few existing studies show that prosodic processing is supported by distributed cortical network
  – which is differentially activated depending on function and frame length (linguistic vs paralinguistic; tone vs intonation; tones vs segments)
  – including structures in STG and LIFG for processing linguistic information


BUT

• These studies do not distinguish between gradient and categorical variation in form and meaning
Prediction

• If hierarchically organised processing is a universal characteristic of language processing, encompassing segmental as well as intonational information

• we should expect to observe dissociations in lower-level auditory and higher-level linguistic subprocesses which reflect distinctions made in linguistic theory
Hypotheses

1. intonation is best analysed in terms of distinct phonological and phonetic levels of representation, which are defined by their contribution to meaning (‘linguistic’ and ‘paralinguistic’), and can be realised through gradient as well as categorical variation in form.

2. the neurobiological processing of intonation is hierarchically organized, reflecting these functional and formal distinctions.
Functional Magnetic Resonance Imaging experiments

- Method: compare activation levels for auditory stimuli
- Systematically varying categorical and gradient properties reflecting intonational form and function
- In a comprehension task
- The experiments will use different levels of background noise to distinguish between more abstract linguistic processing and less specialised acoustic processing by comparing activations for stimuli in different noise conditions

fMRI: Activation measured as differences in the blood oxygenation level-dependent (BOLD) signal
Functional Magnetic Resonance Imaging experiments

- Areas specialised for speech processing sensitive to level of distortion: affects intelligibility. Within those areas:
  - Lower-level acoustic processing of intonation contours distinguished from more abstract linguistic processing when BOLD response varies as function of categorical change in intonation contour (rise vs. fall)
  - Where nature of judgement (linguistic vs paralinguistic meaning, e.g. interrogativity vs. Surprise) identifies subsystems involved
Functional Magnetic Resonance Imaging experiments

• Areas specialised for speech processing sensitive to level of distortion: affects intelligibility. Within those areas:
  – Lower-level acoustic processing of intonation contours distinguished from more abstract linguistic processing when BOLD response varies as function of categorical change in intonation contour (rise vs. fall)
  – where nature of judgement (linguistic vs paralinguistic meaning, e.g. interrogativity vs. Surprise) identifies subsystems involved

• Experiment replicated with gradient change in form of intonation contours (variation in peak height in rises) to examine interaction with gradience in form
Production and perception experiments: Goals

A series of production and perception experiments will explore pitch variation in utterance-final rises and falls in Southern British English to obtain

• independent behavioural/linguistic evidence to support Hypothesis 1, and

• a better understanding of the phonetic detail of the cues involved (duration, loudness, spectral tilt, and their interaction with pitch), which will also inform stimulus development in the fMRI studies.
Combining tasks

Different tasks are combined, because:

- no single task can test categorical and gradient variation in form as well as meaning.
- tried and tested methods in intonational phonology, but they have never been simultaneously applied to the same sets of stimuli before.
Production

Rises, falls and fall-rises are elicited in contexts

• controlled for stress and segmental make-up

• which induce readings in different linguistic and paralinguistic functions:
  – question, statement, and continuation
  – in neutral or surprised contexts

Acoustic measurements of duration, amplitude, spectral tilt and F0 in the accented syllable and the tail of the contour to investigate cue dependencies in the functional contexts
F0 alignment and scaling

- F0 minimum
- F0 maximum

/ y n i m ə s /
Production data

H*L (statement)

L*H (question)

H*H (question)

H*LH

neutral question

H*LH

neutral question

H*LH

neutral question

H*LH

neutral question
Perception: Paradigms
traditionally used to investigate intonational contrast

• **Imitation task** (Pierrehumbert and Steele 1989)
• **Pitch range task** (Gussenhoven and Rietveld 1997)
• **Semantic tasks: interpretation of intonation contour as**
  - rating on semantic scale (Uldall 1964, Grabe et al 1997)
  - yes/no decision about appropriateness of meaning (Bartels and Kingston 1994)
  - judgement of acceptability in context (e.g. Caspers 1998)
• **Categorical Perception paradigm** (Liberman et al. 1957; Studdert-Kennedy et al. 1970)
Examples from research on French intonation

- Intonational grammar of Post (2000, 2002) predicts:
  There are two phonological categories of rises in French:
  rise to high and rise to middle of speaking range

- Where the timing of the start of the rise is gradenitly variable (phonetic): starting point rise

\[T'as\ \textit{vu} \ \textit{Marianne}, \ \textit{Paul}, \, \ldots\ \]
\[T'as\ \textit{vu} \ \textit{Marianne}?\]
Perception experiment

• Question: How many categorical distinctions in final rises?
• Systematic variation of:
  A - peak height
  B - starting point of rise
  C - steepness of rise
• Judgements:
  1. identification (categorisation)
  2. same or different (discrimination)
  3. attitude/emotion (semantic task)

Post (1999)
Semantic task

Phonetically different

Alors, c’est pas vrai?

‘So it isn’t true?’

Phonologically the same

Contrastive judgements
Categorical Perception findings

- Identification and discrimination only match categorical perception in A:
  Only difference in peak height is phonological
- Confirms the hypotheses
- But gradiently varying timing of rise also meaningful (early: surprised)
- and baseline didn’t work...

(cf. Ladd and Morton 1997)
Expected outcomes

• Evaluation of Autosegmental-Metrical tenet of phonetics-phonology dichotomy provides significant contribution to intonational phonology

• Neuroimaging evidence for a hierarchically organised cortical system parallel to that found for segmental variation will represent significant advance in neurolinguistics

• Information about relevance of phonetic and phonological variation in intonation in speech perception can be applied to improve speech synthesis and recognition, language teaching and speech therapy