Categories and gradience in intonation: An fMRI study
Brechtje Post¹, Emmanuel Stamatakis¹, Iwo Bohr*, and Francis Nolan#
¹RCEAL, ¹Anaesthesia, and #Linguistics, University of Cambridge, UK
bmbp2@cam.ac.uk

The Autosegmental-Metrical (AM) framework for intonational analysis [9] is now firmly established as the predominant theoretical framework in the field. The central insight on which the framework builds is that intonation independently carries linguistic meaning which is conveyed by abstract categorical phonological elements that are physically instantiated in a gradient way during phonetic implementation (e.g. [5,7]). However, empirical support for this distinction between phonology and phonetics in intonation has proved elusive (e.g.[8], cf.[5]). The problem is that categories and gradient variation in intonational forms are closely intertwined, since both can in fact be used to convey meaning, e.g. [1,3,11]. For instance, a gradiently wider pitch excursion can be used to signal gradiently increasing surprise (called ‘paralinguistic’ here), but gradient variation in form can also signal categorically distinct meanings, as when a bigger pitch excursion for an utterance-final rise signals a question instead of a continuation (‘linguistic’ here).

In this paper, we investigate the neural substrates for the processing of phonetic as opposed to phonological information in intonation for the first time, combining perception data with direct physical evidence from functional Magnetic Resonance Imaging. The underlying assumption is that the different levels of representation of phonological and phonetic variation in intonation mirror differential activations in a distributed cortical network of hierarchically organised neural subsystems which subserve different cognitive functions in speech comprehension, cf. [2,4,6,10].

Using an event-related design, we recorded BOLD responses in the 3T Siemens Tim Trio MRI scanner at the MRC-CBU (Cambridge, UK) for 15 participants who made linguistic or paralinguistic interpretations of auditory stimuli in a forced choice speeded response task. Using Praat, fundamental frequency (F0) was resynthesised on 24 words with 5 intonation contours (Table 1). This speech condition was replicated as a hummed condition by low-pass filtering the stimuli. The images were re-aligned, spatially normalised, and analysed in SPM8. Two GLM designs at the subject level (one non-parametric, the other with the contours as linear parametric modulators) were carried forward in a random-effects analysis at the group level.

Linguistically interpreted stimuli activated a widespread network of sites including STG bilaterally and LIFG, as we hypothesised, as well as areas that are likely to show activation due to the task (button pressing; Figure 1, left panel). Paralinguistic interpretation engaged the same fronto-temporal network to a lesser extent, but crucially, the activations that were observed for the linguistic and paralinguistic conditions differed as a function of F0 contour. The parametric GLM analysis showed that, in fact, the variation in F0 in the different contours had a significant effect only in the linguistic condition, this effect being observed specifically in areas in middle temporal gyrus bilaterally, in left supramarginal and inferior parietal regions, and in right angular gyrus (Figure 1, right panel); a similar network to that previously found for interrogative rises vs. declarative falls [12]. The categorical perception data mirror these findings, confirming that the interactions between intonational cues in signalling meaning simultaneously depend on F0 contour and intonational function.

Taken together, these findings suggest that a specialised system supports the processing of linguistic phonological information in intonation, as distinct from paralinguistic phonetic information. This would imply that hierarchically organised processing encompasses segmental as well as intonational properties, reflecting distinctions made in current phonological theory, including the AM framework.
<table>
<thead>
<tr>
<th>Linguistic condition</th>
<th>Paralinguistic condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F0 manipulation</strong></td>
<td></td>
</tr>
<tr>
<td>Monotone</td>
<td>24 stimuli</td>
</tr>
<tr>
<td>-3ST fall</td>
<td>24 stimuli</td>
</tr>
<tr>
<td>+3ST rise</td>
<td>24 stimuli</td>
</tr>
<tr>
<td>+6ST rise</td>
<td>24 stimuli</td>
</tr>
<tr>
<td>+9ST rise</td>
<td>24 stimuli</td>
</tr>
<tr>
<td>(same stimuli)</td>
<td>(same stimuli)</td>
</tr>
</tbody>
</table>

Table 1: Experimental design

Figure 1: Non-parametric analysis (left): Network including inferior frontal and temporal regions activated for the linguistic interpretation of a 3ST rise compared to null events; Parametric analysis (right): Activations in angular gyrus and medial temporal gyrus in the linguistic condition as a function of variation in F0. Contrasts thresholded at p<0.001 voxel level, and p<0.05 FWE correction at cluster level.

References