The processing of question-intonation: An fMRI study.

Colin P Doherty [1], W. Caroline West [1], Laura C Redi [2], David Gow Jr[1], Stefanie Shattuck-Hufnagel [2], David Caplan [1]

1. Neuropsychology Lab and Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, USA

2. Speech Communication group, Massachusetts Institute of Technology, Cambridge USA

ABSTRACT

We examined changes in fMRI BOLD signal associated with question/statement judgments to investigate the neural basis of one aspect of intonation. Stimuli consisted of voice recordings of 150 sentences in English and white noise bursts in an event-related paradigm. Sentence types included pairs of lexically identical items, with a falling intonation for sentences (FS) and rising intonation for questions (RQ) and a third stimulus type with a falling intonation and a word order change creating a question (FQ). Functional echo planar imaging (EPI) scans were collected from 11 normal subjects who listened to stimuli and made question/statement judgments with a button press. Results reveal increased BOLD activity in bilateral inferior frontal and temporal regions for RQ over either FQ or FS indicating at the very least a robust processing of acoustic differences. The involvement of left frontal regions not thought to be involved in pitch determination may be a clue to categorical phonological processing. Increased BOLD activity in left lateral temporal regions for questions of any form (RQ or FQ) over FS indicates regions responsive to meaningful (illocutionary) differences between questions and statements.

INTRODUCTION

Intonation refers to the use of suprasegmental phonetic features to convey sentence-level meanings. Intonation may be used to convey categorical linguistic contrasts (such as question/statement judgments) or non-categorical "paralinguistic" contrasts such as emotional states (Ladd, 1996). The act of "posing a question" may be signaled lexically by the use of wh- words, by a change in word order, or prosodically, through the use of intonation (Couper-kuhen, 1986). In this study, we sought to investigate the neural basis of question/statement judgments through the isolated alteration of the intonation contour in a given sentence.

Data from both lesion studies and more recent functional imaging tend to agree that emotional prosody is largely right lateralized (Ross and Mesulam, 1979; Buchanan *et al.*, 2000). Studies looking specifically for the representation of 'linguistic prosody' have sometimes used question-statement judgments and have reported deficits due to damage in either hemisphere suggesting a cooperative role for each side (Heilman *et al.*, 1984). Using auditory stimuli in normal subjects, one recent functional imaging study showed a bilateral and equal increase in BOLD for both phoneme detection and question-statement intonation judgments (Stiller, 1997). The results confirm the notion of an equal and cooperative role for the hemispheres in perceiving linguistic prosody. However, there has been no attempt to tease apart what has been called the illocutionary or semantic/pragmatic role of intonation from syntactic markers in question/statement judgments.

Further predictions regarding the anatomical localization of intonation processing can be made by examining the results of other acoustic, phonological and semantic processing studies of human voice sounds. While pitch processing in tones has been isolated to the right frontal areas (Zatorre, 1994), activation studies of complex human voice stimuli have consistently failed to demonstrate differential activation between the left and right superior temporal gyrus (STG) when speech sounds are compared against silence (Binder *et al.*, 1994). In adjacent superior temporal sulcus (STS), speech comparisons have indicated a leftward asymmetry and a preference for processing words over non-words, indicating that the STS may have a linguistic role at the phonological level (Binder *et al.*, 1997). Comparing words and other perceptual semantic stimuli also activates the inferior frontal gyrus (IFG) (Poldrack *et al.*, 1999). This area has also shown activation in tasks associated with attention to phonemes, indicating a possible role in phonological encoding (Paulescu *et al.*, 1993). The activation off the more lateral and ventral perisylvian areas in tasks of semantic processing have been interpreted as reflecting the processing of non-acoustic elements in speech such as lexical semantic features, a theory that is supported by the presence of naming and semantic deficits in areas outside the STG (; Damasio *et al.*, 1996).

On the basis of the work reviewed above, we hypothesized that the vascular response to manipulation of phonologically relevant intonation would activate a network of language related cortex encompassing basic acoustic processing of voice on both supratemporal planes, pitch discrimination in right inferior frontal regions and sentence level phonological semantic processing in left inferior frontal and temporal areas.

Materials and Methods

Subjects: Eleven normal subjects (seven female; mean age: 23.1 years, range: 18-26; mean years of education: 15.5 years) were recruited. All subjects were right-handed native English speakers.

Stimulus Materials: Stimuli consisted of auditory recordings of a human voice enunciating sentences in English. 150 triads of sentences based on concept stems, were constructed. Each sentence triad consisted of a string of lexical items, which were intoned as (1) a question with a rising boundary tone (RQ -"She was talking to her father?"),(2) a statement with a falling boundary tone (FS – "She was talking to her father") or (3) as a question with a word-order change to denote the illocutionary intent resulting in a question with a falling countour (FQ – "Was she talking to her father?"). Ten naive subjects rated the emotional valence of each sentence on a 7-point scale ranging from highly unemotional to highly emotional. Mean ratings for RQ, FQ and FS were 4.1+/- 0.9, 3.9+/- 0.6 and 4 +/- 0.6 respectively. An analysis of variance (ANOVA) revealed no significant effects of sentence type (F (2, 14) = 1.72, p = 0.24).

Psycholingusitic/fMRI methods: A pair of electrostatic headphones was placed comfortably over both ears. Surrounding this was a tight fitting helmet of neoprene rubber designed to reduce extraneous scanner noise. Each sentence was presented during a 4 second epoch. The three sentence types were presented in random order along with white noise trials of variable duration. During the auditory stimulus presention (sentences and white noise), a white fixation cross on a black background was presented via a projected screen. Subjects made question/statement discrimination judgements with both hands using a custom designed magnet-compatible button box. Stimuli were presented over 5 runs of 240 seconds each.

Magnetic Resonance Imaging Procedure: Each subject participated in two scanning sessions. In the structural session, two sets of anatomical images were acquired in a 1.5T whole-body Siemens Sonata scanner (Siemens Medical Systems, Iselin, NJ) using a T1-weighted MP-RAGE sequence (TR = 7.25 ms, TE = 3.0 ms, and flip angle = 7°). Volumes consisted of 128 sagittal slices (effective thickness = 1.33mm; matrix size = 192 x 256; FOV = 256 mm; in-plane resolution = 1.33 mm x 1.0 mm). The functional session utilized a 3.0T head-only Siemens Allegra scanner. Scans included a lower-resolution T1-weighted MP-RAGE sequence, (TR = 11.08 ms, TE = 4.3 ms, flip angle = 8°) which consisted of 128 sagittal slices (effective slice thickness = 1.33 mm, matrix size = 128 x 256, FOV = 256 mm, in-plane resolution = 2.0 mm x 1.0mm). A T1-weighted inversion-recovery echo-planar sequence (TR = 6s, TE = 29 ms, flip angle = 90°) with 33 slices aligned parallel to the line defined by the anterior- and posterior-commissures (AC-PC) was also acquired to aid registration of the functional images with the anatomical images. The functional volume acquisitions utilized a T2*-weighted gradient-echo pulse sequence (TR = 2 s, TE = 25 ms, and flip angle = 90°). The volume comprised 33 transverse slices aligned along the AC-PC plane. Each run consisted of 120 volume acquisitions for a total of 3960 images.

Cortical Surface Reconstruction: The high-resolution anatomical MP-RAGE scans were used to construct a model of cortical surface. An average of the two structural scans was used to maximize the signal to noise ratio. The cortical reconstruction procedure involved (1) segmentation of the cortical white matter, (2) tessellation of the estimated border between gray and white matter, providing a geometrical representation for the cortical surface of each subject; and (3) inflation of the folded surface tessellation to unfold cortical sulci, allowing visualization of cortical activation in both the gyri and sulci simultaneously (Dale et al., 1999, Fischl et al, 1999a, Fischl *et al.*, 2001). For purposes of inter-subject averaging, the reconstructed surface for each subject was morphed onto an average spherical representation. This non-rigid surface-based deformation procedure results in a substantial reduction in anatomical and functional variability across subjects relative to the more commonly used normalization approaches.

Functional Pre-processing: For each subject, the functional volumes of each participant were spatially smoothed using a three-dimensional technique and a full-width half-max (FWHM) value of 6mm. Normalization to correct for signal intensity changes and temporal drift was then added. The functional images for each subject were co-registered to his or her high-resolution anatomical volume. The selective averages and variances of each subject's functional data were resampled onto his or her inflated cortical surface and into the spherical coordinate system using the surface transforms described above. Each subject's data were then smoothed on the surface tessellation using an iterative nearest-neighbor averaging procedure equivalent to applying a two-dimensional Gaussian smoothing kernel with a FWHM of approximately 8.5 mm.

Statistical analysis: The pre-processed functional data were averaged across subjects on a per-voxel basis using a random effects model. With the end of the auditory sentence occurring at 4 seconds, the vascular response to the auditory perception of sentence-final intonational change would be expected to start by 6 seconds and to be established by 10 seconds. Thus the BOLD signal was collapsed across the 3 TR intervals of 6, 8 and 10. In an exploratory approach to the data, statistical activation maps were constructed for contrasts of interest, and activation coordinates in spherical space that were significant at the p<0.01 level and, that occurred in clusters of five or more significant voxels were retained. When significant peaks occurred within 8mm of one another, the most significant peak was retained. These peak activations represent regions of activation for the comparison. In an region of interest (ROI)-based approach to data analysis, 17 hypothesis driven anatomical ROIs on both sides were defined in

spherical anatomical space corresponding to cortex considered to be involved in acoustic/phonetic, phonological and semantic processing. Graphs of the haemodynamic response were constructed and expressed as a percentage change in BOLD signal. The haemodynamic response was averaged across the same TR intervals of 6, 8 and 10 in each trial type in these voxels to yield a single percentage change value for each subject. The responses were analyzed in an ANOVA with the factors of hemisphere (2), ROI (17) and stimuli (3). Further analysis of significant effects and interactions was performed using Tukey's least mean squares adjustment for multiple comparisons.

RESULTS

I. Accuracy Data: The mean and standard deviation percentage scores for discrimination accuracy of RQ, FQ and FS were 92.8 +/- 5.7%, 91.4 +/- 5.7% and 94.2 +/- 6.3% respectively. An ANOVA revealed no significant main effects of sentence type ($F(2, 27) = 0.51$, $p = 0.61$).
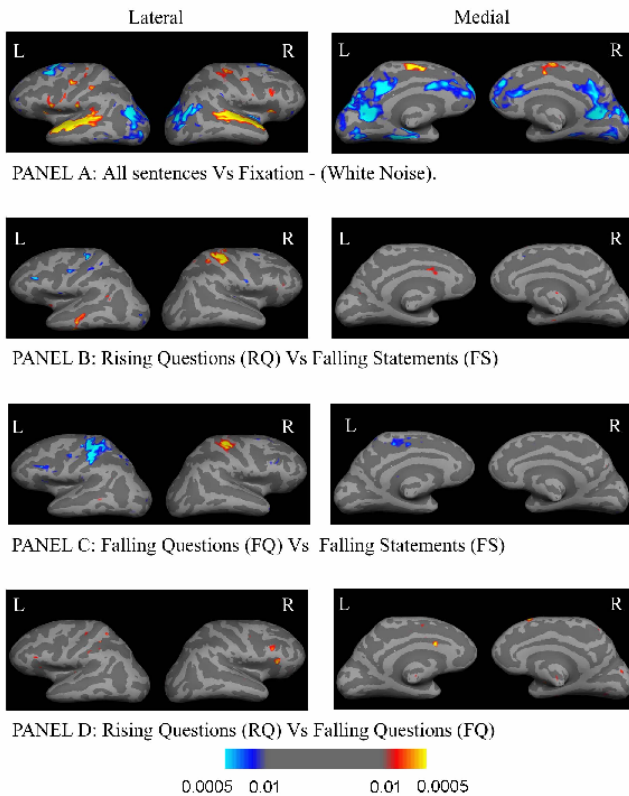
II. Imaging Data

Figure 1. Statistical parametric map of BOLD signal activation in relation to stimulus comparisons.

A. Statistical parametric maps: Figure 1 shows the statistical activation maps of the BOLD signal averaged across eleven subjects superimposed upon an averaged inflated brain. Panel A shows the BOLD activity associated with comparing the activity generated by sentences to the white noise fixation. Increased BOLD signal for spoken sentences over the fixation condition (red/yellow) using a statistical threshold of p<0.01 was seen in both superior temporal regions, and to a lesser extent both lateral inferior frontal regions. There was also increased activation in motor and premotor areas consistent with push button activity. Panels B, C and D show the statistical maps generated by paired comparisons of the BOLD activity in each of the main conditions. Increased and corresponding decreased activity seen in homologous motor areas in the condition comparisons is consistent with hand motion on the right in response to statements and on the left in response to questions. When the BOLD activity generated by Rising Questions (RQ) was compared to Falling Statements (FS) (Panel B), increased activity for RQ over FS was noted in the left lateral temporal lobe. There was also a small amount of activity observed in the IFG bilaterally, greater for FS on the left and RQ on the right. Finally, there was also a focus of increased activity favoring RQ in the area of anterior cingulate cortex. When all the BOLD activity generated by the Falling Questions (FQ) was compared to FS (Panel C), there was increased activity for FS in the left inferior frontal area. Finally, when the BOLD activity generated by RQ was compared to FQ (Panel D), there was increased activity for RQ mainly in the right IFG, but also the anterior cingulate and the left IFG. Other effects that were too subtle to be seen on the voxel-based map were investigated by a ROI approach.

B. Region of interest analysis: BOLD response for each stimulus relative to baseline was analyzed in an ANOVA with the factors of hemisphere (2), ROI (17) and sentence type (3). There were statistically significant main effects of hemisphere ($F(1, 1121) = 5.1$, $p < 0.05$), and ROI ($F(16, 1121) = 13.8$, $p < 0.0001$), and a trend towards significance for sentence type ($F(2, 1121) = 3.22$, $p = 0.0613$). However, these results were qualified by a significant interaction of hemisphere, region and sentence type ($F(32, 1121) = 1.9$, $p < 0.005$). Tukey's test revealed significant differences in 9 regions, which are summarized in Table 1 (below).

Region of Interest	Lt hemisphere % change in BOLD signal	sig (p value)	Rt hemisphere % change in BOLD signal	sig (p value)
Sup Temp Gyrus	RQ>FQ	<0.0001	RQ>FQ	<0.0001
	RQ>FS	0.0013	RQ>FS	*0.065
Sup Temp Sulcus			RQ>FQ	*0.08
			RQ>FS	0.014
Heschl Gyrus	RQ >FQ	<0.0001	RQ>FQ	<0.0001
			RQ>FS	<0.0001
Middle Temporal Gyrus			RQ>FS	0.0001
Circular Sulcus of Insula			RQ>FQ	0.0156
Inf Frontal Gyrus-pars orb			RQ>FQ	<0.0001
			RQ>FS	<0.0001
Inf Frontal Gyrus-pars oper	RQ >FQ	0.023		
Inf Frontal Gyrus -pars tri	RQ >FQ	<0.0001	RQ>FQ	<0.0001
			RQ>FS	0.0012
	FS > FQ	*0.072		
Inf Frontal Sulcus			RQ>FQ	0.0006
Pre-central Gyrus	FQ>FS	0.0003		

Table 1: Significant (p<0.05) (*trend) stimulus comparisons derived from Tukey's least mean squares adjustment in 9 of the 17 hypothesized regions.

The results of the ROI analysis can be summarized as follows: 1) There was a significant increase in BOLD activity for RQ over both FQ and FS in a number of right-sided ROIs (three temporal regions - the STG, Heschl's gyrus, the STS and two inferior frontal areas (pars orbitalis and pars opercularis) and in one left sided region, the STG. 2) There were a number of ROIs on both sides that demonstrated an increase in BOLD signal for RQ over FQ only. Two were on the right (the circular sulcus of the insula and the inferior frontal sulcus) and three were on the left; Heschl's gyrus and the pars opercularis and triangularis of the IFG. 3) One region, the pars triangularis in the left IFG, demonstrated a trend towards increased activity for FS over FQ.

DISCUSSION

This study demonstrates the sensitivity of event-related fMRI to hemodynamic changes in regions of the brain involved in processing of question/statement judgements using phonologically specified intonational contrasts. Sentences were compared which (1) differed in both pitch contour and meaning (illocution) (RQ vs. FS); (2) had a common pitch contour thus isolating illocutionary force (FQ Vs. FS) and (3) had common illocutionary value and different pitch contours. (RQ Vs. FQ).

The most robust finding in this study was that rising questions with a high ending boundary tone caused preferential activation in primary and secondary association perisylvian regions and inferior frontal areas bilaterally compared to phrases that have a falling boundary tone. These results indicate areas that are, at the very least, responding to the acoustic changes in intonation. The question remains, however, given the linguistic role we have proposed for intonation, whether we can identify areas where phonological rather than purely acoustic processing takes place. It could be argued that the activation of the left inferior fontal areas in this study by sentences with a rising boundary tone implies a level of processing beyond the mere acoustic. Evidence for this argument is that prior activation studies have isolated lexical phonological encoding to this region, which has been interpreted as a "sensory-motor phonetic encoding" area (Preslau *et al*., 1993).

Exploring semantic differences between questions and statements; the statistical activation maps showed increased activation for both RQ and to a lesser extent FQ in the left lateral temporal areas (middle and inferior temporal cortex) when compared to FS. This suggests the processing of meaning rather than simply an acoustic/phonetic difference. However, when comparing FQ to FS there was a robust increase in activation, this time favoring FS (appearing as blue on the statistical map), in the left IFG. Similarly, the ROI analysis showed a trend towards significance for increased activation of FS over FQ in the pars triangularis of the left IFG. With both boundary tones falling in this comparison, the activation must be related to meaningful differences rather than acoustic between the stimuli.

CONCLUSION

These experiments are a first step in determining the neural substrate of intonation as a linguistic operation. However, it is not possible on the basis of this data alone to disentangle the acoustic from the linguistic with these stimuli, which were intended as a first-pass exploration of the neural responses to intonation. Further experimentation examining the neural response to acoustic differences of equal magnitude that fall within and between illocutionary categories could further specify the role of different areas identified here in intonational processing.

REFERENCES

[1] Binder J, Rao S, Hammeke T, et al. Functional magnetic resonance imaging of human auditory cortex. *Ann Neurol* **35**, pp. 662-72, 1994.

[2] Buchanan T, Lutz K, Mirzazade S, Specht K, Shah N, Zilles K *et al*. Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cogn Brain Res* **9**, pp. 227-38, 2000.

[3] Couper-kuhen E. An introduction to English prosody. London: Edward Arnold. 1986.

[4] Dale A, Fischl B, Sereno M. Cortical surface-based analysis - I. Segmentation and surface reconstruction. *Neuroimage* **9**, pp. 179-94, 1999

[5] Damasio H, Grabowski T, Tranel D, Hichwa R, Damasio A. A neural basis for lexical retrieval. *Nature* **380**, pp. 499-505, 1996.

[6] Fischl B, Liu A, Dale AM. Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* **20**, pp. 70-80, 2001.

[7] Fischl B, Sereno M, Dale A. Cortical surface-based analysis II Inflation, flattening, and surface-based coordinate system. *Neuroimage* **9**, pp195-207, 1999a.

[8] Heilman K, Bowers D, Speedie L, Coslet H. Comprehension of affective and non-affective prosody. Neurology **34**, pp. 917-21, 1984.

[9] Heilman K, Scholes R. Auditory affective agnosia. JNNP **38**: pp. 69-72, 1975

[10] Ladd D. Intonational phonology. Cambridge: Cambridge University Press., 1996.

[11] Paulescu E, Frith C, Frackowiak R. The neural correlates of the verbal component of working memory. *Nature* **362**(6418), pp. 342-5, 1993.

[12] Ross E, Mesulam M. Dominant language functions in the right hemisphere: Prosody and emotional gesturing. *Arch Neurol* **36**, pp. 144-8, 1979

[13] Stiller D, Gaschler-Markefski B, Baumgart F, Schindler F, Tempelman C, Heinze H *et al*. Lateralized processing of speech prosodies in the temporal cortex: A 3-T functional magnetic resonance imaging study. *MAGMA* **5**, pp. 275-84, 1997.

[14] Zatorre, R. Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* **14**, pp. 1908-19, 1994.