Categories and gradience in intonation

Citation for published version:
Post, B, Stamatakis, E, Bohr, I, Nolan, F & Cummins, C 2015, Categories and gradience in intonation: A functional Magnetic Resonance Imaging study. in J Romero & M Riera (eds), The Phonetics-Phonology Interface: Representations and methodologies. John Benjamins, pp. 259-284. DOI: 10.1075/cilt.335

Digital Object Identifier (DOI):
10.1075/cilt.335

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
The Phonetics-Phonology Interface

Categories and gradience in intonation

A functional Magnetic Resonance Imaging study[*]

Brechtje Post, Emmanuel A. Stamatakis, Iwo Bohr,
Francis Nolan & Chris Cummins
University of Cambridge

The Autosegmental-Metrical framework (AM) assumes that a distinction needs to be made between linguistic phonological information (categorical) and paralinguistic phonetic information (gradient) in intonation. However, empirical evidence supporting this assumption has proved to be elusive so far. In this study we analysed whether the theoretical distinction is reflected in perceptual biases and neural activation in the brain. The results of a combined behavioural and neuroimaging study demonstrate that intonational function indeed activates different but overlapping neural networks with more widespread activation for categorical phonological stimuli, especially in middle temporal gyrus bilaterally and left supramarginal and inferior parietal areas. In contrast, for paralinguistic gradient stimuli activation is restricted to right inferior frontal gyrus. These neural differences mirror differences in response times in a listening experiment testing categorical perception for the same stimuli. These findings support a theoretical model of intonation, such as AM, in which linguistic and paralinguistic information are distinguished.

1. Introduction

Intonation is notoriously difficult to analyse because it is carried by a continuous sound signal, it has multiple functions, and it interacts with other elements in the speech signal that convey meaning. We do know, however, that at some stage in the comprehension process, some of the continuous information is interpreted categorically and decoded into

[*] This research was funded by an ESRC First Grant (RES-061-25-0347; PI Post, CI Stamatakis). We are very grateful to two anonymous reviewers for their very helpful comments on the paper, to the radiographers and administrative staff at the MRC Cognition and Brain Sciences Unit in Cambridge for their support with data collection, to Toby Hudson for his assistance with the development of the design and materials for the experiment, and to Elaine Schmidt for her help with editing.

distinct meaningful units, such as a rising pitch pattern that can be used to mark an interrogative, as opposed to a falling pattern for a declarative, as is illustrated in Table 1 (a). Here, the sentence-level pragmatic meaning is affected in a categorical way (also 'linguistic meaning'; Ladd 1996). The intonational information can also make a more gradient contribution to meaning, when gradual increases or decreases in a particular feature like pitch convey, for instance, a more angry or less timid tone of voice. In such cases, the emotional or attitudinal meaning of the message is affected ('paralinguistic' meaning), as in Table 1 (d). These variations in form and their contribution to meaning are closely intertwined, and difficult to disentangle. One reason is that both categorical and gradient variation in form can in fact map on to categorical linguistic as well as gradient paralinguistic variation in meaning, shown in Table 1 (b) and (c), respectively (e.g., Crystal 1969, Bolinger 1970, Scherer, Ladd & Silverman 1984; cf. Taylor 2003).[1] Table 1 (b) shows an example of a categorical difference in form signalling a difference in paralinguistic meaning. Here, a rising pattern indicates an interrogative interpretation of an utterance, while a fall-rise in the same context could make the interrogative sound surprised (depending on dialect and other contextual factors). Conversely, the varying height of the final rise in Table 1 (c) can be associated with a categorical distinction in linguistic meaning, for instance when a bigger pitch excursion signals that the speaker is asking a question instead of holding the floor with a continuation rise.[2]

---

[1] We concentrate on pitch (F0) here, since it is generally assumed to be the primary correlate of intonation (e.g., Bolinger 1986, Cruttenden 1986, Gussenhoven 2004). Evidently, other parameters like loudness, duration and voice quality are also at issue (Post et al. 2007 for an overview), and will therefore have to be controlled for in any experiments. Rises are selected because they are relatively well-understood and have long been the focus of the debate on intonational meaning (Ladd 1981).

[2] In Table 1, categorical differences in form are contrasts in pitch direction (e.g., rise vs. fall), and categorical differences in meaning are taken to be differences that affect the linguistic message in a categorical fashion (e.g., question vs. statement). This classification simplifies the complexities in form-meaning relations in prosody, since, for

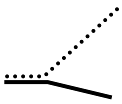
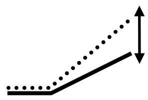
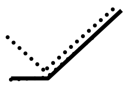
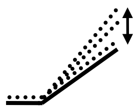
		Form	
		Categorical	Gradient
Meaning	Categorical	(a) Linguistic:  e.g., declarative fall vs. interrogative rise	(c) Linguistic:  e.g., continuation vs. interrogative rise (where the latter could have a higher peak than the former)
	Gradient	(b) Paralinguistic:  e.g., ‘neutral’ interrogative rise vs. surprised interrogative fall-rise	(d) Paralinguistic:  e.g., interrogative rise expressing various degrees of surprise (i.e., peak height varies with surprise)

Table 1: *Categorical and gradient variation in form and meaning in intonation, with stylised fundamental frequency contours illustrating differences in form*

The Autosegmental-Metrical (AM) framework for intonational analysis (Pierrehumbert 1980, Gussenhoven 1984, 2004, Ladd 1996, Jun 2005) has proved to be an excellent vehicle for disentangling the complexities of the relation between form and meaning, as well as allowing us to model how they are to some extent intertwined (cf. Post, D’Imperio & Gussenhoven 2007). This is because it crucially distinguishes between, on the one hand, abstract phonological (categorical, discrete) representations which independently carry linguistic meaning, and on the other, the phonetic implementation of those representations in speech production and perception (e.g., Ladd 1996, Gussenhoven 2004). For

instance, the paralinguistic meaning of a message can also contrast categorically (e.g., a speaker either does or does not sound surprised). The crucial difference that we are emphasising here is that ‘unsurprised’ and ‘surprised’ are the end-points of a continuum of increasing surprise, while ‘statement’ and ‘question’ contrast paradigmatically.

instance, the H*L and the L*H pitch accent (i.e., a fall and a rise) are categorically different forms in Southern British English which are used to signal categorically different meanings, such as the declarative vs. interrogative contrast illustrated in Table 1 (a). Their actual phonetic realisation depends on speaker characteristics and context. Thus, most women tend to produce wider pitch excursions than men, but excursions are also wider in speech produced in noise (Shriberg et al. 1996). Conversely, pitch excursions may be smaller than usual when there is little scope for voicing in the segmental material (e.g., Grabe et al. 2000). This type of phonetic variation is systematic and gradient, and does not affect meaning. Note, however, that phonetic variation can be exploited in the formation of phonological categories in L1 acquisition (Best, this volume), which implies that this kind of intonational variation could also be used in early perceptual attunement in a similar way.

In the Autosegmental-Metrical framework, linguistically structured phonological information is also distinguished from paralinguistic information, which is iconic, and largely independent of the individual language (Gussenhoven 2004). Thus, when the size of a pitch excursion signals paralinguistic meaning, as is exemplified in Table 1 (d), this variation does not affect the core linguistic message. Paralinguistic information is placed outside the phonology, and accounted for in the phonetics. Cases like those sketched in Table 1 (c) would be accounted for in the phonology if they signalled a difference in the tonal configuration that specifies the two types of rise. For instance, LH*H% could be the tonal representation for the interrogative and LH*0% for the continuation (as in e.g., Post 2000 for French; 0% represents an Intonation Phrase boundary which is not specified for tone). The difference in F0 scaling would be a result of the phonetic implementation of the two tonal configurations. In this example, the sequence of two high tones in the interrogative would result in a realisation ending in a higher peak than the configuration with only one high tone (in French, both tones would normally be realised on the phrase-final syllable). By distinguishing between phonological categories and phonetic realisation in this way, the framework allows us to make clear hypotheses about how exactly acoustic variation maps onto meaning in intonation. Since the framework offers discrete, economical, insightful and — crucially — testable formalisations of intonation systems, it promises to provide a key to understanding cross-linguistic and stylistic variation in intonation patterns, and their role in language processing and the neural architecture that supports it.

However, although the Autosegmental-Metrical framework is now firmly established as the predominant theoretical framework in the field, empirical support for this distinction between phonology and phonetics has proved elusive (e.g., Ladd & Morton 1997; cf. Gussenhoven 2004). Very few empirical studies have attempted to disentangle the interactions in form and meaning at issue here (but see e.g., Post 2000, Chen 2005), not least because little is known about the phonetic detail of the cues involved (Post, D'Imperio & Gussenhoven 2007). To date, cognitive, neuropsychological and neurobiological studies of prosody — in which such interactions are a common confound — have not addressed the issue at all. Inevitably, this has led to widely diverging conclusions about the neural underpinnings of prosody (summarised in e.g., Mayer et al. 2002).

In this paper, we will provide direct evidence from neuroimaging testing the phonetics/phonology distinction that underpins the Autosegmental-Metrical framework. This approach rests on the assumption that the different levels of linguistic representation of categorical (phonological) and gradient (phonetic) intonation mirror differential activations in a distributed cortical network of hierarchically organised neural subsystems which subserve speech comprehension (cf. Coleman 1998, Haspelmath 2004, Indefrey & Levelt 2004, Poldrack 2006). Focussing on the contrast between categorical linguistic and gradient paralinguistic intonational information (i.e., Table 1(a) and (d); the most common form-function pairings within the four-way schema), we expect to observe differential neural processing which would not only serve to support the theoretical distinction that is made between phonological categories versus phonetic realisation in intonation in Autosegmental-Metrical theory (cf. Eulitz & Lahiri 2004 for segmental structure), but it would also allow us to pin down the neural architecture that supports the processing of the intonational information.

1.1 The neural basis of prosodic processing

Originating in neuropsychological studies, hypotheses about prosodic processing in the brain centre around either a stimulus-dependent interpretation, with neurobiological specialisation for specific aspects of the acoustic signal such as duration, pitch and intensity (e.g., Zatorre et al. 1992, Mayer et al. 2002, Gandour et al. 2003a), or a task-dependent or domain-dependent interpretation in which speech has a unique neural substrate, and different functional properties of speech are subserved by different mechanisms, such as linguistic prosody by left hemisphere mechanisms, and affective or emotional prosody by right hemisphere ones (van Lancker 1980, Wildgruber et al. 2004; cf. Schirmer & Kotz 2006, Mitchell & Ross 2008, Leitman et al. 2010).

Gandour and colleagues have explored the neural correlates of linguistic and paralinguistic prosody in a series of fMRI experiments in which they contrasted lexical tone in Chinese with a range of other prosodic phenomena (Gandour et al. 2003a, 2003b, 2003c, 2004; cf. Krishnan, Gandour & Bidelman 2010). In the first study, they compared intonation and emotion, and found that when linguistic interpretation of the stimuli was required, the frontoparietal region in the left hemisphere was preferentially activated, whereas emotional prosody preferentially activated the same region on the right (also, Grandjean & Scherer 2006). The second study showed that lexical tone, which has a short frame length (i.e., extending over a single syllable), preferentially activates frontoparietal regions in the left hemisphere, while intonation, with its longer frame length (extending over the phrase), activates frontoparietal regions bilaterally, even when the actual duration of the stimuli is the same. Elsewhere, tonal contrasts have been found to elicit larger MMN responses when listeners are exposed to native tonal contrasts (Chandrasekaran, Krishnan & Gandour 2009, Ren, Yang & Li 2009), and when the tonal stimuli cross a category boundary (Chandrasekaran, Krishnan & Gandour 2007, Xi et al. 2010). In the third study by Gandour and colleagues, activation was shown to be modulated as a function of the subsyllabic unit involved (tones, rhymes or consonants). Within left inferior prefrontal cortex, posterior/dorsal regions are implicated in the extraction of phonologically relevant information (both segmental and suprasegmental), and these subregions are functionally distinct from anterior/ventral regions in left inferior prefrontal cortex which are activated by attention to semantic properties. The fourth study confirmed that, when acoustic or auditory features are related to conceptual (linguistic) representations, the perception of prosody becomes lateralised to task-dependent regions in the left hemisphere. Results by Wildgruber et al. (2004) comparing linguistic with paralinguistic intonation, and by Doherty et al. (2004) comparing declarative falls with interrogative rises have further confirmed that distinct inferior frontal and temporal regions are involved in the processing of intonation depending on the communicative function of the cues involved (cf. Borràs Comes et al. 2012).

These findings point toward a distributed cortical network underlying prosodic processing which is differentially activated depending on communicative function, where linguistic intonation is supported by structures in left inferior frontal gyrus and bilateral superior temporal gyrus, while paralinguistic function tends to be right-dominant, but other factors like frame length also play a role. Unfortunately, it is difficult to relate the findings directly to our research question, since gradient and categorical variation in form and meaning are routinely confounded in previous studies.[3] This implies that we cannot disentangle the contribution of differences in form and meaning to the patterns of activation that have been observed, and as a consequence, we cannot pinpoint the neural substrate for abstraction and categorisation in the processing of linguistic information (i.e., the phonology) as distinct from gradient paralinguistic information (part of the phonetics).

The study reported here is more comprehensive than existing studies in that it takes the interaction between gradience and categories in form and meaning into account, and asks to what extent linguistic intonational information is encoded in a way that is comparable to other types of abstract categorical information in speech (section 1.2).

1.2 Neural processing hierarchies for abstraction and categorisation of speech sound

Current models of speech processing in the brain paint a complex picture of multiple processing streams involving anatomically separable areas that are interconnected through multiple pathways, and which support several distinct levels of processing serially and in parallel (e.g., Hall et al. 2002, Davis & Johnsrude 2007). Initial processing of the incoming speech signal in neocortex takes place in auditory cortex bilaterally, with different subfields showing selective responsiveness to different spectro-temporal properties (e.g., Obleser, Lahiri & Eulitz 2004). From auditory cortex, hierarchical connections between auditory core, belt and parabelt areas project to distributed, interconnected fields in superior temporal gyrus

[3] With the exception of Borràs Comes et al. (2012), who used ERPs to examine cases like (c) and (d) in Table 1.

(STG) and sulcus (STS), the inferior parietal lobule and in prefrontal cortex. This cortical system supports at least three, and possibly four, discrete levels of auditory processing (e.g., Kaas, Hackett & Tramo 1999, Tramo et al. 2005; cf. McLachlan & Wilson 2010; Obleser & Eisner 2009 for a recent review), and two distinct, functionally specialised parallel processing streams can be distinguished within this network: a dorsal and a ventral stream (Hickok & Poeppel 2000, 2007, Scott & Johnsrude 2003). For abstraction and categorisation in speech, it has been claimed, interactions with the dorsal-stream network ensure that 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 (Davis & Johnsrude 2007).

A number of linguistically informed studies have confirmed that processing of contrastive segmental information is hierarchically organised and tends to involve the dorsal-stream network sketched above, including structures in STG and left inferior frontal gyrus (LIFG; Burton, Small & Blumstein 2000, Boatman 2004, Eulitz & Lahiri 2004, Obleser, Lahiri & Eulitz 2004), together with supramarginal gyrus for stimuli representing a phonological change. (Dehaene-Lambertz et al. 2005, Obleser et al. 2006). Obleser, Lahiri & Eulitz (2004), for instance, succeeded in dissociating activation for different place features that encode contrastive segmental information, and Eulitz & Lahiri (2004) conducted a study in which they distinguished between the processing of underlying phonological representations and surface phonetic forms (cf. MEG and EEG findings in Dehaene-Lambertz 1997, Phillips et al. 2000, and Sharma & Dorman 2000). Phonological and morpho-phonological processing have also been shown to be dissociable in a fronto-temporal network linking anterior cingulate, LIFG and bilateral STG (Tyler et al. 2005).

These findings show that speech input that functions contrastively is treated differently at the neural level. Categorical phonological distinctions that are made in linguistic theory are found to have distinct neural correlates, with preferential activation for (morpho)phonological information in superior temporal and frontal areas which are not engaged when ‘low-level’ acoustic information is being processed.

1.3 Hypothesis

The neurobiological processing of intonation is hierarchically organised in a distributed cortical network including the temporo-parietal-frontal areas which are typically recruited in speech processing more generally. Within this network, linguistic (phonological) intonation preferentially activates left hemisphere structures that support higher-level phonological speech processing (in particular LIFG, STG/MTG). Paralinguistic (phonetic) intonation is more strongly right-lateralised.

2. Methodology

2.1 Design

A functional Magnetic Resonance Imaging (fMRI) experiment evaluated brain activations elicited in a comprehension task by utterances with different kinds of rises and falls (see section 2.2 for stimulus details; activation levels were measured as differences in the blood oxygenation level-dependent (BOLD) signal). Rises and falls were chosen for the intonational form condition (Table 2) because they appear to have elicited wider activation maps in the two previous fMRI studies that tested similar conditions (Doherty et al. 2004, Wildgruber et al. 2004). The intonational form condition (5 levels) was fully crossed with an intonational function condition (2 levels: interrogativity for linguistic meaning vs. surprise for paralinguistic meaning) to disentangle categoricity and gradience, and both conditions were replicated as unintelligible hummed stimuli in a fully matched control condition (‘Hum’ in Table 2; see 2.2 for stimulus generation).

The rationale for this design was that, if speech processing is hierarchically organised in the brain, speech-specific processes should be distinguishable from less specialised acoustic processes, and for speech-specific processes, higher-level phonological abstraction should be distinguishable from lower-level phonetic decoding (cf. Davis & Johnsrude 2003). The former can be identified by comparing activations for real speech stimuli and stimuli which are speech-like, but unintelligible (hum). Thus, the hummed signal will generate an elevated BOLD response in all areas that are recruited for processing auditory input, including areas that are specialised for speech processing, as opposed to the speech stimuli in the experiment (‘Speech (Words)’ in Table 2), which will generate differential activation in speech-specific areas only.

		Experimental: Speech (Words)		Control: Hum	
		Linguistic condition	Paralinguistic condition	Linguistic condition	Paralinguistic condition
F_0 manipulation	+9ST rise	24 stimuli	(same stimuli)	24 stimuli	(same stimuli)
	+6ST rise	24 stimuli	(same stimuli)	24 stimuli	(same stimuli)
	+3ST rise	24 stimuli	(same stimuli)	24 stimuli	(same stimuli)
	Monotone	24 stimuli	(same stimuli)	24 stimuli	(same stimuli)
	–3ST fall	24 stimuli	(same stimuli)	24 stimuli	(same stimuli)
Baseline control		Rest: 60 null events			

Table 2: *Experimental design (F_0: fundamental frequency, ST: semitone)*

Within speech-specific areas, we will be able to identify the two neural subsystems that are involved in the processing of linguistic phonological and paralinguistic phonetic intonation by examining areas that show an elevated BOLD response to stimuli when they are interpreted for their linguistic as opposed to paralinguistic meaning, as well as the overlap between those areas of activation. Cutting across the function conditions, the intonational form condition ('F0 manipulation' in Table 2) can be used to distinguish lower-level acoustic processing of intonation contours from more abstract linguistic processing when areas in which the BOLD response varies as a function of a categorical change in intonation contour (rise vs. fall) are compared with those in which it varies more gradiently (different F0 peak heights in a rise).

A resting baseline served as a second control condition ('null events' in Table 2), which was used to increase design efficiency, and to validate the experimental set-up by verifying whether the activation maps for general auditory and speech-specific stimuli were as expected. Here, participants were asked to focus on a cross-hair that was centred on the screen, without auditory input. The analysis (subtraction: all auditory stimuli minus all null events; not reported below) showed the expected activations for auditory input in bilateral temporal areas responsible for auditory processing, including higher-level auditory/speech areas on the left, as well as primary motor and higher-order senso-motor areas consistent with right hand button pressing.

2.2 Stimuli

Using Praat (Boersma & Weenink 2010), fundamental frequency was resynthesised on 24 items to create 5 different intonation contours, as in Figure 1. A comparison between responses to steps 1-5 allows us to test for the effect of intonational function (linguistic vs. paralinguistic) when a categorical difference in form is involved (i.e., categorical form + categorical meaning vs. categorical form + gradient meaning), while a second comparison for steps 3-5 allows us to test for the effect while excluding the potential confound of the categorical distinction in form (e.g., related to the fact that the falls in our stimuli are less likely to express interrogativity than surprise; i.e., gradient form + categorical meaning vs. gradient form + gradient meaning).

The items were bi- or trisyllabic geographical place names with initial or penultimate stress, selected for their sonorance to facilitate F_0 tracking and F_0 manipulation during resynthesis (e.g., *Manila*, *Angola*, *Uganda*). Place names were chosen so as to ensure that the stimuli were semantically neutral and unmarked for affect and interrogativity (either morpho-syntactically or pragmatically; cf. Gandour et al. 2003b). The items were digitally recorded in the sound-proof booth of the University of Cambridge Phonetics Laboratory at 48 KHz by a male native speaker of standard Southern British English with a background in phonetics. The items that were used for resynthesis were realised with a single falling accent which was produced with a narrow pitch range, resulting in nearly monotonous utterances.

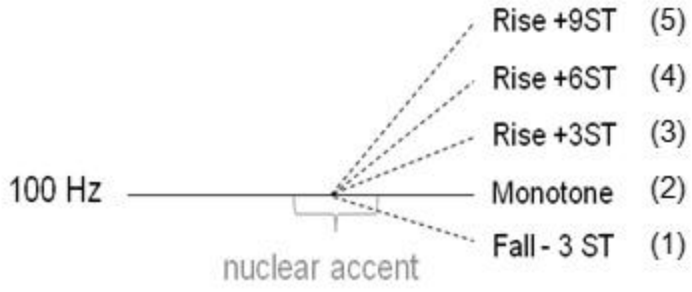


Figure 1: *Diagrammatic representation of the F_0 manipulations in the experimental stimuli*

A monotone base was created from each recorded item by equalising intensity to 80 dB, and F_0 contour to 100 Hz, approximating the mean values of the original recordings. Four stimuli were resynthesised from each base by varying the F_0 slope from the accented syllable to the end of the word in steps of 3 semitones (see Figure 1), creating one falling movement and three rising movements. The starting point of the F_0 movement in the accented syllable was determined by hand on the basis of a second set of recordings of the same words in which the speaker had been asked to produce falls and rises; this F_0 turning point usually coincided with the offset of the accented vowel. The resynthesized stimuli were evaluated for their naturalness and valence (interrogativity and surprise).

The resulting 120 speech stimuli were transformed into their unintelligible counterparts by low-pass filtering (stop Han band with range 300-12000 Hz and smoothing frequency 100 Hz). This resulted in a set of stimuli which sounded like vocal hums, but which retained the F_0, intensity, and durational characteristics of the speech stimuli.[4]

2.3 Participants and procedure

Using an event-related design, we recorded BOLD responses in the 3 Tesla Siemens Tim Trio MRI scanner at the MRC Cognition and Brain Sciences Unit (Cambridge, UK). Fifteen right-handed native speakers of standard Southern British English (mean age 23, 9 women) without neurological or psychiatric disorders, head injury, or hearing impairment took part in the experiment.

In two separate blocks of 300 events each (60 null events, 120 speech stimuli, and 120 ‘hum’ stimuli; 11.5 minutes per session), participants were cued to make a forced choice speeded identification response,

[4] A pilot behavioural study evaluated stimulus quality and their valance for the intonational functions used in the experiment (interrogativity and surprise), using a semantic rating task (Uldall 1964). The results confirmed that the stimulus types were interpreted differently for the two functions.

evaluating the interrogativity or the surprise signalled by the stimulus. We opted for an explicit task rather than passive listening so as to ensure that participants processed the F_0 variations in the signal for the intended meanings. The responses were elicited in two separate sessions (linguistic and paralinguistic) for the same reason. Seven of the participants started with the linguistic block, and eight with the paralinguistic block.

Each block contained all 120 speech stimuli, all 120 hums, and 60 null events, presented in a pseudo-random order with maximally three consecutive stimuli of the same type (F_0 contour or speech/hum), and in which any series of eight non-null events was interspersed with two non-consecutive null events. Each event lasted 2.3s, consisting of 0.5s silence, 0.4-0.6s stimulus presentation (depending on stimulus length), and 1.8s time out measured from stimulus offset (Figure 2). The response was cued by a question which was displayed on a screen for the entire duration of the session (*Does this sound like a statement?* for the linguistic condition, and *Does this sound surprised?* for the paralinguistic condition), except during null events, when the text was replaced by a fixation cross. Participants were familiarised with the task in a two-minute practice session outside the scanner. A high-fidelity stimulus delivery system was used for stimulus presentation and to record button-presses, using E-Prime v1 (Professional Psychology Software Tools, Inc. Pittsburgh, PA) with Etymotic ER-3 headphones and an MRI-compatible button box.

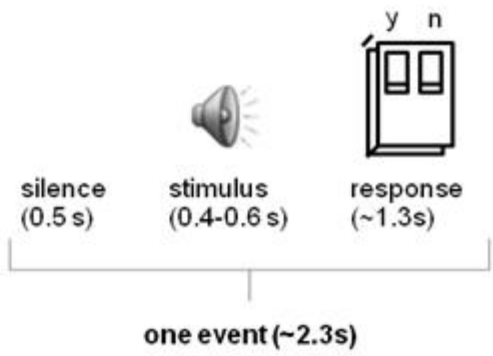


Figure 2: *Schematic representation of an experimental event*

Images were acquired with an EPI T2* sequence with TR=2s, TE=30ms, flip angle=78°, fov=192mm x 192mm, resolution=3×3mm, and a 3mm gap between slices with 32 oblique axial slices per volume.

2.4 Image analysis

The images were re-aligned, spatially normalised to a standard EPI template (based on the Montreal Neurological Institute (MNI) reference brain), smoothed with a full-width half-maximum 6 mm isotropic Gaussian kernel, and statistically modelled in SPM8 (Statistical Parametric Mapping, Friston et al. 1997) implemented in Matlab 7. Two General Linear Model (GLM) designs were implemented at the subject level, one non-parametric, and the other with the F_0 contours as linear parametric modulators. The subject-level designs were carried forward in group-level random-effects analyses. Subtraction analyses at the group level based on the non-parametric design were used to explore differential brain activation in the intonational form and the intonational function conditions (cf. Henson 2006), while the parametric design allowed us to directly contrast activations for linguistic and paralinguistic meaning independent of the effect of variation in form when the variation in form is gradient (i.e., F_0 variation in the three rises, here). In other words, a linguistic functional distinction for rising movements could be contrasted with a paralinguistic function for the same rises, excluding the potential confound of the categorical distinction in form (fall vs. rise).

The prediction would therefore be that, while variation in form has an effect that is independent of function, the same forms elicit different patterns of neural activation depending on the communicative function of the intonation contour, where (a) linguistic intonation predominantly activates a network including areas previously observed for other ‘phonological’ processing (LIFG, STG/MTG), and (b) paralinguistic intonation is more strongly right lateralised.

Group level contrasts were thresholded at $p<0.001$ at a voxel level, and only clusters that survived $p<0.05$ FWE correction for multiple comparisons are reported. The Automated Anatomical Labelling (AAL) toolbox for SPM was used to name the anatomical areas where peak activity voxels were located (Tzourio-Mazoyer et al. 2002). Brodmann areas were identified using the Talairach Daemon software (Lancaster et al. 1997, 2000), based on coordinate values converted from MNI to Talairach using a non-linear transformation as implemented in the mni2tal.m Matlab script (Brett et al. 2001); see Appendix for table of activations.

3. Results

The first subsection presents the behavioural results for the identification task. In the second subsection, the results of the subtraction analyses of brain activation in the experimental conditions are reported (i.e., the non-parametric design). In the third subsection, only rising stimuli are included in a set of analyses which explore the difference between linguistic and paralinguistic function when variation in form is factored out (i.e., the parametric design).

3.1 Behavioural data

The raw identification data, plotted in Figure 3 (top panel), were transformed for statistical analysis in order to factor out the task-related difference between the linguistic and paralinguistic condition (i.e., in the linguistic condition, 'yes' responses are associated with the lowest step of the F_0 continuum, but in the paralinguistic, they are 'no' responses).

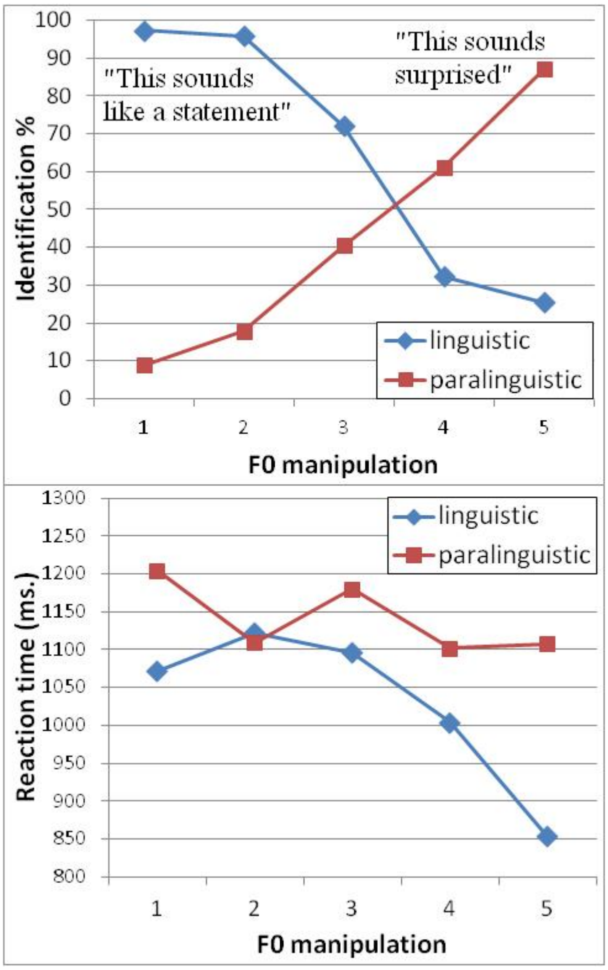


Figure 3: *Identification function (top panel) and mean reaction time (bottom panel) for the 5 F_0 manipulations (1 = 3ST fall, 2 = monotone, 3 = 3ST rise, 4 = 6ST rise, 5 = 9ST rise) in the linguistic and paralinguistic conditions*

In the identification data (Figure 3 top panel), the effects of the intonational form condition (F_0 manipulation) were significant for both intonational functions (linear mixed effects model regression analysis implemented in R; coefficient = −0.400, Z = −17.8, p < 0.001), indicating that the responses varied significantly as a function of F_0 contour. The effect for intonational function was also significant across F_0 steps (coefficient = −0.474, Z = −3.19, p = 0.0014), and the interaction between F_0 step and intonational function was nearly significant (coefficient = 0.0807, Z = 1.90, p = 0.058). This indicates that the linguistic and paralinguistic condition yielded significantly different identification response patterns over the different steps of the F_0 continuum tested in the experiment. The data in the linguistic condition showed a non-linear response curve with an abrupt transition between majority responses — a shape typically associated with more categorical perception. The data in the paralinguistic condition displayed a nearly perfectly linear response curve — a shape predicted under gradient perception.

The reaction times that were associated with the identification responses (Figure 3 bottom panel) also showed significant effects for intonational form (coefficient = –3.83, t = –1.99, p < 0.05) and function (coefficient = 47.6, t = 3.74, p < 0.001), with higher F_0 steps leading to faster decisions overall, and with paralinguistic interpretations yielding longer latencies overall than linguistic interpretations. A further analysis showed that the response latencies for stimuli that straddled the borderline between the hypothesised categories were responded to significantly more slowly than could be expected, all else being equal (i.e., steps 3 and 4 with 30%-75% identification rates in Figure 3 top panel; coefficient = 38.0, t =

2.64, $p < 0.01$; intonational form and function remained significant in this analysis).[5]

3.2 fMRI data: Non-parametric design

A baseline subtraction analysis which was carried out to identify activation for intelligible speech in the experiment ('words') as distinct from speech-like auditory input ('hums') revealed activation of the speech processing system typically observed for auditory linguistic experiments involving higher-order phonological processing of speech, comprising large areas of activation in bilateral auditory temporal areas, as well as clusters of activation in left inferior frontal cortex overlapping with Broca's area, in the left cerebellum, and in the right putamen (Figure 4 panel A; see Appendix for table of activations).

Two further subtraction analyses explored brain activation in the linguistic and paralinguistic conditions for intelligible speech while abstracting away from general auditory processing of speech-like input with the same form properties (primarily F0 manipulation, here). These analyses revealed widespread activations in superior and medial temporal gyrus bilaterally for both conditions (Figure 4 panels B and C), but with more activation in the linguistic condition especially in the left hemisphere, extending further to the anterior and posterior regions of the superior and middle temporal gyri, and including left inferior frontal gyrus, perisylvian cortical areas, and parietal regions as well as the putamen (Figure 4 panel B). In the paralinguistic condition, activations were restricted to superior temporal gyrus bilaterally (Figure 4 panel C). A paired t-test analysis of the

---

[5] The effect of 'borderline-ness' is obscured in Figure 1 by the main effect of intonational form, which speeds up responses in the linguistic condition such that rises have shorter reaction times than falls. However, if form alone determined response times, a linear response curve would be predicted; the 'borderline-ness' effect is visible in the significantly slower RTs at steps 3 and 4 than a linear response curve would predict.

first level contrasts for the linguistic and paralinguistic conditions (for words>hums) revealed significant differences in temporal areas only (including right superior temporal and left medial temporal gyrus; Figure 4 panel D).

A final subtraction analysis examined activations associated with the form condition by contrasting falls and rises directly (–3ST fall - +9ST rise), showing large clusters of activations in postcentral and middle temporal areas bilaterally, as well as right hemispheric activations in the supplemental motor area and supramarginal gyrus (not shown; see Appendix for details).

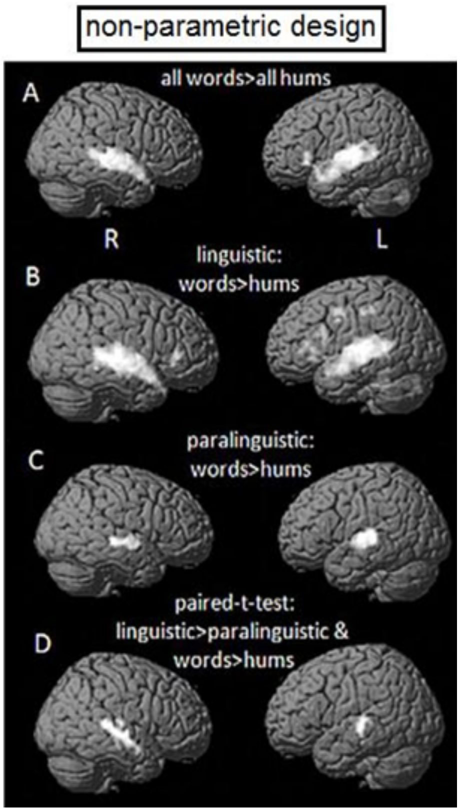


Figure 4: *Non-parametric design: hemodynamic responses specific for (A) intelligible speech, (B) a linguistic interpretation, (C) a paralinguistic interpretation, (D) a linguistic versus paralinguistic interpretation*

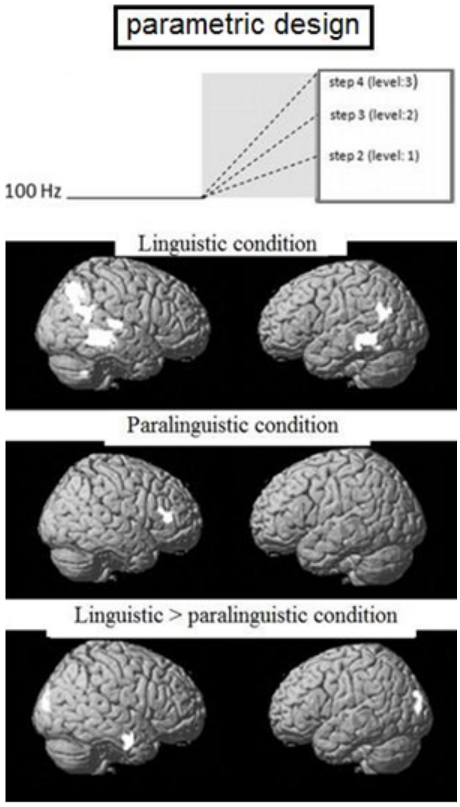


Figure 5: *Parametric design: hemodynamic responses in the linguistic condition (top), paralinguistic condition (centre), linguistic vs. paralinguistic condition (bottom); $p < 0.001$ at voxel level, FWE correction at $p < 0.05$ for cluster level*

3.3 fMRI data: Parametric design

A second set of subtraction analyses based on a parametric design was carried out to contrast activations for linguistic and paralinguistic meaning while excluding the potential confound of the categorical distinction in form (i.e., fall versus rise in the intonational form condition; Figure 5). By including 'gradient' variation in the intonational form condition as linear parametric modulators (i.e., F_0 manipulation in the shape of the 3 levels of rise in Table 1), the effect of a categorical phonological functional distinction for rising movements could be contrasted with a paralinguistic ‘phonetic’ function for the same rises.

As before, a wider network of activations was observed for the linguistic condition than the paralinguistic condition, with bilateral middle temporal and right superior temporal activations, and parietal regions encompassing, on the left, an area at the interface between the temporal and parietal lobe and in supramarginal gyrus, and on the right, angular gyrus, as well as a small cluster in the cerebellum (Figure 5 top panel; see Appendix for table of activations). In the paralinguistic condition, only right hemispheric activations were found in inferior frontal gyrus (Figure 5 centre panel). A direct contrast between the two function conditions revealed differential activation in the left cuneus and the right inferior temporal gyrus extending to middle temporal gyrus (Figure 5 bottom panel).

4. Discussion

In our experiment, linguistically interpreted stimuli activated a widespread network of sites including, as we hypothesised, superior and medial temporal areas bilaterally as well as a small cluster in left inferior frontal gyrus overlapping with Broca’s area — brain structures implicated in higher order phonological processing of speech processing more generally (e.g., Burton, Small & Blumstein 2000, Gandour et al. 2003a, 2003c, 2004, Eulitz & Lahiri 2004, Obleser, Lahiri & Eulitz 2004, Zhang et al. 2011; reviews in Obleser & Eisner 2009 and Price 2010). In addition, they activated an area in left cerebellum which is often observed in tasks involving the processing of prosody and words (e.g., Binder et al. 2009).

When we factored out the effect of differences in form (F0 contour) in the parametric GLM analysis, linguistic interpretation engaged areas in middle temporal gyrus bilaterally, left supramarginal and inferior parietal regions, and right angular gyrus, delineating a network that is similar to that previously found for linguistic interpretation of interrogative rises as opposed to declarative falls (Doherty et al. 2004). Within this network, the middle temporal gyrus is considered to be active at an intermediate level of speech processing, in between "lower-level" audition involving core and parabelt areas in auditory cortex, and "higher-level" combinatory processes in comprehension which engage areas in (pre)frontal cortex. Therefore, prefrontal areas are a likely candidate for top-down modulation of the activity in the medial temporal gyrus for linguistic intonation (cf. Doherty et al. 2004), in line with proposals by Davis & Johnsrude (2007) for abstraction and categorisation in neural speech processing more generally. The activity in the supramarginal gyrus could reflect access to already-abstracted higher-level phonological information, which has also been observed in other studies in which it was associated with the processing of a phonological change, but not with acoustic differences (Obleser & Eisner 2009).

Paralinguistic interpretation engaged the same fronto-temporal network to a lesser extent, but activations were only right-dominant when variation in form was factored out in the parametric analysis. Here, activation was restricted to the right inferior frontal gyrus, which has often been shown to be implicated in the processing of emotional prosody (Schirmer & Kotz 2006).

Directly contrasting the two functions confirmed that linguistic and paralinguistic intonation are differentially processed, even when the same forms are used as stimulus material. However, contrary to our assumption, activations only differed significantly in the temporal lobe, and not elsewhere in the network that is recruited for phonological processing. The absence of a significant effect in left inferior frontal areas which we expected to find could be due to an interaction between form and function, similar to that revealed in the behavioural data.

Taken together, these findings suggest that a specialised system supports the processing of linguistic phonological information in intonation as distinct from paralinguistic phonetic information, and that, within this system, the processing of intonational function interacts with intonational form. The processing of linguistic intonational information recruits the same neural systems and mechanisms that support abstraction and categorisation in speech more generally, contrary to what is often assumed in the literature. We also observed very similar dissociations in lower-level auditory and higher-level linguistic subprocesses, and we observed interactions with areas that are known to process already-abstracted phonological information exclusively in the linguistic condition. This suggests that the system is hierarchically organised, and that interactions with the dorsal-stream network ensure abstraction and categorisation (Davis & Johnsrude 2007). The behavioural data also support this interpretation, since they confirmed that the interactions between intonational cues in signalling meaning simultaneously depend on F0 contour and on communicative function (linguistic or paralinguistic). Here, responses in the linguistic condition were compatible with categorical perception, while those in the paralinguistic condition were typical for continua that are perceived gradiently. This implies that the distinction between phonetics and phonology which is made in linguistic theory (e.g., the Autosegmental-Metrical framework) is reflected in the neural architecture that supports the processing of intonational information.

Since the processing of linguistic and paralinguistic meaning engages two heavily overlapping networks which show clear but quite small clusters of differential activation, and since intonational form and meaning appear to interact in determining patterns of neural activation, it would be interesting to explore to what extent time course differences in patterns of activation rather than localisation per se are key in the neural mechanisms at play here.

5. Conclusion

Intonational function plays a crucial role in the neural processing of speech prosody, where different but overlapping cortical networks in both hemispheres contribute differentially to the processing of different intonational functions. In addition, the processing of linguistic information was found to resemble that of other categorical phonological information in the speech signal. This finding can be interpreted to support theoretical models of intonation in which linguistic and paralinguistic information are crucially distinguished, as in the Autosegmental-Metrical framework for intonation analysis. This implies that hierarchically organised neural processing encompasses suprasegmental (prosodic) as well as segmental properties, and hence, that it may well be a universal characteristic of language processing.

References

Best, Catherine T. 2015. "Devil or Angel in the Details?: Perceiving phonetic variation as information about phonological structure". This volume.

Binder, Jeffrey R., Rutvik H. Desai, William W. Graves & Lisa L. Conant. 2009. "Where is the Semantic System?: A critical review and meta-analysis of 120 functional neuroimaging studies". *Cerebral Cortex* 19:12.2767-2796.

Boatman, Dana. 2004. "Cortical Bases of Speech Perception: Evidence from functional lesion studies". *Cognition* 92:1-2.47-65.

Boersma, Paul & David Weenink. 2010. *Praat: Doing phonetics by computer: Version 5.2.* (http://www.praat.org/) (Retrieved on October 29, 2010).

Bolinger, Dwight. 1970. "Relative Height". *Prosodic Feature Analysis* ed. by Pierre R. Léon., Georges Faure & André Rigault, 109-125. Paris: Didier.

Bolinger, Dwight. 1986. *Intonation and its Parts: Melody in spoken English*. Stanford, Calif.: Stanford University Press.

Borràs-Comes, Joan, Jordi Costa-Faidella, Pilar Prieto & Carles Escera. 2012. "Specific Neural Traces for Intonational Discourse Categories as Revealed by Human-Evoked Potentials". *Journal of Cognitive Neuroscience* 24:4.843-853.

Brett, Matthew, Alexander P. Leff, Chris Rorden & John Ashburner. 2001. "Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking". *NeuroImage* 14:2.486-500.

Burton, Martha W., Steven L. Small & Sheila E. Blumstein. 2000. "The Role of Segmentation in Phonological Processing: An fMRI investigation". *Journal of Cognitive Neuroscience* 12:4.679-690.

Chandrasekaran, Bharath, Ananthanarayan Krishnan & Jackson T. Gandour. 2007. "Mismatch Negativity to Pitch Contours is Influenced by Language Experience". *Brain Research* 1128.148-156.

Chandrasekaran, Bharath, Ananthanarayan Krishnan & Jackson T. Gandour. 2009. "Relative Influence of Musical and Linguistic Experience on Early Cortical Processing of Pitch Contours". *Brain and Language* 108:1.1-9.

Chen, Aoju. 2005. *Universal and Language-Specific Perception of Paralinguistic Intonational Meaning*. Ph.D. dissertation, Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).

Coleman, John. 1998. "Cognitive Reality and the Phonological Lexicon: A review". *Journal of Neurolinguistics* 11:3.295-320.

Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.

Crystal, David. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Davis, Matthew H. & Ingrid S. Johnsrude. 2003. "Hierarchical Processing in Spoken Language Comprehension". *Journal of Neuroscience* 23:8.3423-3431.

Davis, Matthew H. & Ingrid S. Johnsrude. 2007. "Hearing Speech Sounds: Top-down influences on the interface between audition and speech perception". *Hearing Research,* 229:1-2.132-147.

Dehaene-Lambertz, Ghislaine. 1997. "Electrophysiological Correlates of Categorical Phoneme Perception in Adults". *NeuroReport* 8:4.919-924.

Dehaene-Lambertz, Ghislaine, Christophe Pallier, Willy Serniclaes, Liliane Sprenger-Charolles, Antoinette Jobert & Stanislas Dehaene. 2005. "Neural Correlates of Switching from Auditory to Speech Perception". *NeuroImage* 24:1.21-33.

Doherty, Colin P., W. Caroline West, Laura C. Dilley, Stefanie Shattuck-Hufnagel & David Caplan. 2004. "Question/statement judgments: an fMRI study of intonation processing". *Human Brain Mapping* 23:2.85-98.

Eulitz, Carsten & Aditi Lahiri. 2004. "Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition". *Journal of Cognitive Neuroscience* 16:4.577-583.

Friston, Karl. 1997. "Imaging Cognitive Anatomy". *Trends in Cognitive Science* 1:1.21-27.

Gandour, Jack, Mario Dzemidzic, Donald Wong, Mark Lowe, Yunxia Tong, Li Hsieh, Nakarin Satthamnuwong & Joseph Lurito. 2003a. "Temporal Integration of Speech Prosody is Shaped by Language Experience: An fMRI study". *Brain and Language* 84:4.318-336.

Gandour, Jack, Donald Wong, Mario Dzemidzic, Mark Lowe, Yunxia Tong & Xiaojian Li. 2003b. "A Cross-Linguistic fMRI Study of Perception of Intonation and Emotion in Chinese". *Human Brain Mapping* 18:3.149-157.

Gandour, Jack, Yisheng Xu, Donald Wong, Mario Dzemidzic, Mark Lowe, Xiaojian Li & Yunxia Tong. 2003c. "Neural Correlates of Segmental and Tonal Information in Speech Perception". *Human Brain Mapping* 20:4.185-200.

Gandour, Jackson, Yunxia Tong, Donald Wong, Thomas Talavage, Mario Dzemidzic, Yisheng Xu, Xiaojian Li & Mark Lowe. 2004. "Hemispheric Roles in the Perception of Speech Prosody". *NeuroImage* 23:1.344-357.

Grabe, Esther, Brechtje Post, Francis Nolan & Kimberley Farrar. 2000. "Pitch Accent Realisation in Four Varieties of British English". *Journal of Phonetics* 28:2.161-185.

Grandjean, Didier & Klaus R. Scherer. 2006. "Examining the Neural Mechanisms Involved in the Affective and Pragmatic Coding of Prosody". *Proceedings of the 3rd International Conference on Speech Prosody, Dresden, 2-5 May 2006* ed. by Rüdiger Hoffmann & Hansjörg Mixdorff, paper 268. Dresden: TUD Press.

Gussenhoven, Carlos. 1984. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hall, Deborah A., Ingrid S. Johnsrude, Mark P. Haggard, Alan R. Palmer, Michael A. Akeroyd & A. Quentin Summerfield. 2002. "Spectral and Temporal Processing in Human Auditory Cortex". *Cerebral Cortex* 12:2.140-149.

Haspelmath, Martin. 2004. "Does Linguistic Explanation Presuppose Linguistic Description?". *Studies in Language* 28:3.554-579.

Henson, Richard. 2006. "Forward Inference Using Functional Neuroimaging: Dissociations versus associations". *Trends in Cognitive Sciences* 10:2.64-69.

Hickok, Gregory & David Poeppel. 2000. "Towards a Functional Neuroanatomy of Speech Perception". *Trends in Cognitive Sciences* 4:4.131-138.

Hickok, Gregory & David Poeppel. 2007. "The Cortical Organization of Speech Processing". *Nature Reviews Neuroscience* 8:5.393-402.

Indefrey, Peter & Willem J. M. Levelt. 2004. "The Spatial and Temporal Signatures of Word Production Components". *Cognition* 92:1-2.101-144.

Jun, Sun-Ah, ed. 2005. *Prosodic Typology*. Oxford: Oxford University Press.

Kaas, Jon H., Troy A. Hackett & Mark Jude Tramo. 1999. "Auditory Processing in Primate Cerebral Cortex". *Current Opinion in Neurobiology* 9:2.164-170.

Krishnan, Ananthanarayan, Jackson T. Gandour & Gavin M. Bidelman. 2010. "The Effects of Tone Language Experience on Pitch Processing in the Brainstem". *Journal of Neurolinguistics* 23:1.81-95.

Ladd, D. Robert. 1981. "A First Look at the Semantics and Pragmatics of Negative Questions and Tag Questions". *Proceedings of Chicago Linguistics Society (CLS)*, vol. 17, 164-171. Chicago, Ill.: Chicago Linguistics Society.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert & Rachel Morton. 1997. “The Perception of Intonational Emphasis: Continuous or categorical?”. *Journal of Phonetics* 25:3.313-342.

Lancaster, Jack L., Lacy H. Rainey, J. L. Summerlin, Catarina S. Freitas, Peter T. Fox, Alan C. Evans, Arthur W. Toga & John C. Mazziotta. 1997. “Automated Labeling of the Human Brain: A preliminary report on the development and evaluation of a forward-transform method”. *Human Brain Mapping* 5:4.238-242.

Lancaster, Jack L., Marty G. Woldorff, Lawrence M. Parsons, Mario Liotti, Catarina S. Freitas, Lacy Rainey, Peter V. Kochunov, Dan Nickerson, Shawn A. Mikiten & Peter T. Fox. 2000. “Automated Talairach Atlas Labels for Functional Brain Mapping”. *Human Brain Mapping* 10:3.120-131.

Leitman, David I., Daniel H. Wolf, J. Daniel Ragland, Petri Laukka, James Loughead, Jeffrey N. Valdez, Daniel C. Javitt, Bruce I. Turetsky & Ruben C. Gur. 2010. “It’s not what you Say, but how you Say it: A reciprocal temporo-frontal network for affective prosody”. *Frontiers in Human Neuroscience* 4:19.1-13.

Mayer, Jörg, Dirk Wildgruber, Axel Riecker, Grzegorz Dogil, Hermann Ackermann & Wolfgang Grodd. 2002. “Prosody Production and Perception: Converging evidence from fMRI studies”. *Proceedings of the 1st International Conference on Speech Prosody, Aix-en-Provence, 11-13 April 2002* ed. by Bernard Bel & Isabelle Marlien, 487-490. Aix-en-Provence: Université de Provence, Laboratoire Parole et Langage.

McLachlan, Neil & Sarah Wilson. 2010. “The Central Role of Recognition in Auditory Perception: A neurobiological model”. *Psychological Review* 117:1.175-196.

Mitchell, Rachel L. C. & Elliott D. Ross. 2008. “fMRI Evidence for the Effect of Verbal Complexity on Lateralisation of the Neural Response Associated with Decoding Prosodic Emotion”. *Neuropsychologia* 46:12.2880-2887.

Obleser, Jonas, Henning Boecker, Alexander Drzezga, Bernhard Haslinger, Andreas Hennenlotter, Michael Roettinger, Carsten Eulitz & Josef P. Rauschecker. 2006. “Vowel Sound Extraction in Anterior Superior Temporal Cortex”. *Human Brain Mapping* 27:7.562-571.

Obleser, Jonas & Frank Eisner. 2009. “Pre-Lexical Abstraction of Speech in the Auditory Cortex”. *Trends in Cognitive Sciences* 13:1.14-19.

Obleser, Jonas, Aditi Lahiri & Carsten Eulitz. 2004. “Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Wowels”. *Journal of Cognitive Neuroscience* 16:1.31-39.

Phillips, Colin, Thomas Pellathy, Alec Marantz, Elron Yellin, Kenneth Wexler, David Poeppel, Martha McGinnis & Timothy Roberts. 2000. “Auditory Cortex Accesses Phonological Categories: An MEG mismatch study”. *Journal of Cognitive Neuroscience* 12:6.1038-1055.

Pierrehumbert, Janet. 1980. *The Phonetics and Phonology of English Intonation*. Ph.D. dissertation, Massachusetts Institute of Technology.

Poldrack, Russell A. 2006. “Can Cognitive Processes Be Inferred from Neuroimaging Data?”. *Trends in Cognitive Sciences* 10:2.59-63.

Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*. Ph.D. dissertation, The Hague: Holland Academic Graphics.

Post, Brechtje., Mariapaola D'Imperio & Carlos Gussenhoven. 2007. "Fine Phonetic Detail and Intonational Meaning". *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS16), Saarbrücken, 6-10 August 2007* ed. by Jurgen Trouvain & William J. Barry, 191-196. Saarbrücken: Universität des Saarlandes.

Price, Cathy J. 2010. "The Anatomy of Language: A review of 100 fMRI studies published in 2009". *Annals of the New York Academy of Sciences* 1191.62-88.

Ren, Guiqin. Q., Yufang. Yang & Xiaoqing. Li. 2009. "Early Cortical Processing of Linguistic Pitch Patterns as Revealed by the Mismatch Negativity". *Neuroscience* 162:1.87-95.

Scherer, Klaus R., D. Robert Ladd & Kim E. A. Silverman. 1984. "Vocal Cues to Speaker Affect: Testing two models". *Journal of the Acoustical Society of America* 76:5.1346-1356.

Schirmer, Annett & Sonja A. Kotz. 2006. "Beyond the Right Hemisphere: Brain mechanisms mediating vocal emotional processing". *Trends in Cognitive Sciences* 10:1.24-30.

Scott, Sophie K. & Ingrid S. Johnsrude 2003. "The Neuroanatomical and Functional Organisation of Speech Perception". *Trends in Neurosciences* 26:2.100-107.

Sharma, Anu & Michael F. Dorman. 2000. "Neurophysiologic Correlates of Cross-Language Phonetic Perception". *Journal of the Acoustical Society of America,* 107:5.2697-2703.

Shriberg, Elizabeth, D. Robert Ladd, Jacques Terken & Andreas Stolcke. 1996. "Modeling Intra-Speaker Pitch Range Variation: Predicting f0 targets when 'speaking up'". *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96), Philadelphia, Pa., 3-6 October 1996* ed. by H. Timothy Bunnell, Alfred I. Delaware & William Idsardi, vol. 2, 650-653. Philadelphia, Pa.: Institute of Electrical and Electronics Engineers.

Taylor, John R. 2003. *Linguistic Categorization* (3rd ed.). Oxford: Oxford University Press.

Tramo, Mark. J., Peter A. Cariani, Christine K. Koh, Nikos Makris & Louis D. Braida. 2005. "Neurophysiology and Neuroanatomy of Pitch Perception: Auditory cortex". *Annals of the New York Academy of Sciences* 1060.148-174.

Tyler, Lorraine K., Emmanuel A. Stamatakis, Brechtje Post, Billi Randall & William Marslen-Wilson. 2005. "Temporal and Frontal Systems in Speech Comprehension: An fMRI study of past tense processing". *Neuropsychologia* 43:13.1963-1974.

Tzourio-Mazoyer, Nathalie, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer & Marc Joliot. 2002. "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain". *NeuroImage* 15:1.273-289.

Uldall, Elizabeth. 1964. "Dimensions of Meaning in Intonation". *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday* ed. by David Abercrombie, 271-279. London: Longman.

van Lancker, Diana. 1980. “Cerebral Lateralization of Pitch Cues in the Linguistic Signal”. *International Journal of Human Communication* 13:2.201-277.

Wildgruber, Dirk, Ingo Hertrich, Axel Riecker, Michael Erb, Silke Anders, W. Wolfgang Grodd & Hermann Ackermann. 2004. “Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation”. *Cerebral Cortex* 14:12.1384-1389.

Xi, Jie, Linjun Zhang, Hua Shu, Yang Zhang & Ping Li. 2010. “Categorical Perception of Lexical Tones in Chinese Revealed by Mismatch Negativity”. *Neuroscience* 170:1.223-231.

Zatorre, Robert J., Alan C. Evans, Ernst Meyer & Albert Gjedde. 1992. “Lateralization of Phonetic and Pitch Discrimination in Speech Processing”. *Science* 256.846-849.

Zhang, Linjun, Jie Xi, Guoqing Xu, Hua Shu, Xiaoyi Wang & Ping Li. 2011. “Cortical dynamics of acoustic and phonological processing in speech perception”. *PLoS One* 6:6.e20963.

Appendix

1. Table of activations (peaks of activated clusters): Non-parametric model

	Left hemisphere			Right hemisphere		
Anatomical area (BA)	coordinates (MNI) x y z	size (voxels)	t-value	coordinates (MNI) x y z	size (voxels)	t-value
1. all words>all hums						
superior temporal gyrus/sulcus (21/22)	−60 −16 4	2261	10.03	58 −4 −2	1673	11.27
cerebellum_7b	−16 −78 −44	191	5.61			
insula/inferior frontal gyrus (13/47)	−34 24 2	269	4.64			
putamen	−22 8 −4	269	4.64	24 16 4	178	5.54
2. words>hums in linguistic condition						
superior temporal gyrus/sulcus (21/22)	−58 −26 2	2711	10.84	66 −14 0	2468	11.33
insula	−32 22 4	325	4.22			
fusiform gyrus (36)	−40 −42 −24	176	4.51			
precentral gyrus (6)	−48 0 36	111	5.91			
inferior parietal gyrus (40)	−54 −24 44	100	4.16			
inferior frontal gyrus; *pars triangularis* (45/46)	−48 16 24	198	5.66	54 34 6	107	5.20
putamen	−20 6 −4	185	4.96			
crus of cerebellum	−20−76−36	110	4.80			
caudate				24 22 4	270	7.29
3. words>hums in paralinguistic condition						
superior temporal gyrus/sulcus (21/22)	−60 −14 4	636	6.95	66 −14 −2	288	6.32
4. paired t-test linguistic & paralinguistic words>hums						
superior temporal gyrus				64 −30 6	512	6.33
middle temporal gyrus	−60 −24 0	165	5.23			
5. all words step4>all words step1						
postcentral (2) *extending to inferior parietal*	−44 −30 56	503	7.77	48 −30 54	198	8.03
parietal inferior (*40*) *extending to AG*	−48 −50 38	99	7.24			
posterior portion of middle temporal (21)	−60 −42 −8	134	5.53	52 −28 −8	403	4.33
middle cingulate (32)				4 −18 30	129	4.48
supplemental motor area (6)				2 −4 54	400	4.44
supramarginal gyrus				64 −48 30	377	7.65

2. Table of activations (peaks of activated clusters): Parametric model

	Left hemisphere			Right hemisphere		
Anatomical area (BA)	coordinates (MNI) x y z	size (voxels)	t-value	coordinates (MNI) x y z	size (voxels)	t-value
1. linguistic condition						
posterior portion of middle temporal (21)	−62 −40 −6/ −52 −58 16 −58 −42 −4	304/146 212	7.58/4.91 8.29	54 −46 −4	470	8.54
inferior parietal lobule (40)	−54 −40 44	609	8.58			
supramarginal gyrus (40)	−54 −44 34	354	5.81			
cerebellum 8				20 −56 −44	104	9.88
superior temporal gyrus (40)				54 −20 10	123	5.69
middle occipital gyrus *extending to AG*				36 −64 24	528	5.49
2. paralinguistic condition						
inferior frontal gyrus *pars trigeminalis* (46)				40 36 12	88	5.89
posterior portion of middle temporal gyrus (21)				54 −40 0	609	8.58
angular gyrus/parietal inferior lobule (40)				52 −58 42	266	5.58
3. paired t-test linguistic>paralinguistic						
cuneus	−12 −92 14	177	5.81			
inferior temporal *extending to MTG*				56 −6 −28	88	4.03