Neural correlates of categorical linguistic and gradient paralinguistic intonation

Brechtje Post[*] and Kai Alter[+#]

[*]Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge; [+]Institute of Neuroscience, Newcastle University; [#]Oxford University, Department of Linguistics, Philology and Phonetics
bmbp2@cam.ac.uk, kai.alter@ncl.ac.uk

ABSTRACT

Multiple cues interact to signal multiple functions in intonation simultaneously which makes intonation notoriously complex to analyze. The Autosegmental-Metrical model for intonation analysis has proved to be an excellent vehicle for separating the components, but evidence for the phonetics/phonology dichotomy on which it hinges has proved elusive.

Advocating a multidisciplinary approach, this paper presents ERP evidence that different types of intonational information – linguistic/phonological and paralinguistic/phonetic – recruit overlapping but distinct neural systems, which differ not only in their neural architecture, but also in the time-course of activation in the subcomponents of the systems.

We argue that the findings can be accounted for in a model in which linguistic (phonological) intonation engages a language-specific fronto-temporal system which is specialised for processing categorical linguistic information, while paralinguistic intonation, which reflects biological imperatives more directly, engages a distributed bilateral system which supports perceptual and cognitive processing more generally.

Keywords: Intonational phonology, intonational meaning, pitch, ERP, neural correlates, categories

1. INTRODUCTION

One of the key tenets which underpins the currently predominant theoretical framework for intonation analysis – the Autosegmental-Metrical approach (AM) – is that there is a phonology separate from the phonetics in intonation [28, 19]. In AM, intonation independently carries linguistic meaning (i.e. informational rather than affective or attitudinal meaning) which is conveyed by abstract phonological elements which are physically instantiated during phonetic implementation. The phonological elements are categorical and discrete. For instance, the H*L and the L*H pitch accent (i.e. a fall and a rise) are categorically different phonological forms in Southern British English which are used to signal categorically different meanings (e.g. declarative vs. interrogative). Their actual phonetic realisation depends on speaker characteristics and context. For instance, most women tend to produce wider pitch excursions than men. Excursions are also typically wider in speech produced in noise [30]. Conversely, they may be smaller than usual when there is little scope for voicing in the segmental material, e.g. [14]. This type of phonetic variation is systematic and gradient, and does not affect linguistic meaning.

Distinguishing between phonetics and phonology in this way allows us to draw up testable hypotheses about the precise contribution of intonation in communication. However, incontrovertible evidence to support the tenet itself has remained elusive, e.g. [19, 22]. One reason is that intonation carries paralinguistic as well as linguistic meaning, and the cues to the two types of meaning are difficult to tease apart in the speech signal [5, 9]. For instance, the wider pitch excursion mentioned above can also be used to signal arousal (e.g. anger or excitement), with excursion size correlating directly with degree of arousal. This type of gradient variation in an intonation contour is also meaningful, but since it is not part of the linguistic code, it should be placed in the phonetics ('paralinguistic' [19]).

The multiplicity of relations between form and meaning in intonation has impeded the development of a comprehensive theory of intonational meaning (but see [19]), and obscured evidence for the AM phonetics/phonology distinction. In this paper, we turn to Event-Related Potentials (ERP) in EEG (electroencephalography) to provide direct evidence from neurobiology to support it.

1.1. Neural correlates of intonation

Neurolinguistic studies of intonation have drawn widely diverging conclusions about the neural underpinnings of intonation [23], with a divide between accounts that take a stimulus-based interpretation with neurobiological specialisation for specific acoustic parameters operating over different time-domains, e.g. [15, 20, 29], as opposed to task-based interpretations with neurobiological specialisation for different prosodic functions, e.g. [16, 17, 18, 25]. There are a number of reasons why these studies appear to come to such diverging conclusions. Other than differences in experimental

paradigms and populations, the studies operationalise intonation in very different ways, usually ignoring the complexities of the factors involved in determining form and meaning in intonation sketched above, thus introducing confounds (e.g. contrasting linguistic and paralinguistic meaning while inadvertently covarying form).

In [1] we examined the neural processing of intonational information in a functional Magnetic Resonance Imaging experiment in which we controlled for variation in form while contrasting linguistic and paralinguistic meaning. As we hypothesised, two overlapping neural systems were activated, including superior and medial temporal areas bilaterally [10, 24, 25, 29, 30], as well as a small cluster in left inferior frontal gyrus [10] for linguistic intonation – brain structures implicated in higher order phonological processing of speech processing more generally, e.g. [8, 9, 17, 26, 33], cf. [6]– but right inferior frontal activation for paralinguistic intonation [10, 15], cf. [32]. We also found that activations in the paralinguistic condition were generally weaker (cf. N400 [21], MMN [6]).

One possible explanation is that linguistic and paralinguistic intonation differ in the time course of activation of the overlapping subcomponents of the systems, and not just in the localisation of activation. We explored this possibility in an experiment using ERP, since it provides an excellent tool for examining the latency and amplitude of activation peaks in the relevant neural subcomponents.

1.2. Hypothesis

Categorical linguistic and gradient paralinguistic intonation differentially engage two distinct but overlapping bilateral fronto-temporal neural networks in terms of both latency and amplitude of activation in individual components of the system.

2. METHODS

EEG was recorded while participants performed a categorical perception discrimination task in 20% of the trials, listening to auditory stimuli over headphones.

2.1. Design and stimuli

Categorical and gradient intonational variation were crossed in a 2x2 design, as shown in Fig. 1. The stimuli were single word utterances (neutral place names) with different pitch contours created by means of resynthesis in Praat [4]: (A) fall of 3 semitones (ST, (B) fall of 9 ST, (C) rise of 3 ST, (D) rise of 9 ST. The stimuli were a subset of the stimuli

that were used in the fMRI experiment [1], representing the tops and bottoms of the acoustic continua tested there. In line with the behavioural findings obtained for these same stimuli there, the difference in pitch direction (A&B vs. C&D) was expected to be processed primarily as a categorical linguistic distinction (signalling e.g. question vs. statement in the absence of communicative context), whereas the difference in pitch excursion is more likely to be interpreted as a gradiently varying paralinguistic difference here (A&B vs. B&D; signalling e.g. different levels of arousal when angry or excited).
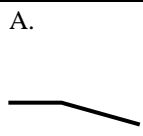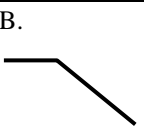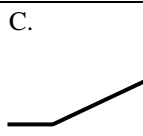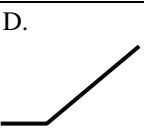

Figure 1: Experimental design

2.2. Participants

21 Right- handed participants between the ages of 17 and 32 were recruited. Participants were given a questionnaire to complete asking for basic personal information and information about neurological and mental health. Of the 10 participants whose data were included in the final analyses, 6 were female. None reported any neurological disorders.

2.3. Procedure

Using a forced choice speeded response discrimination task, participants were asked to detect any difference in pitch contour between the stimuli within each pair (same/difference response). During testing, participants stared at a fixation cross on a computer screen while pairs of stimuli were presented over headphones: a baseline stimulus (A) followed by either another A or one of the deviant stimuli B, C or D. The participants were told to pay attention to the pairs of stimuli and detect any difference in pitch contour between the stimuli within each pair. Following 20% of the stimulus pairs, the participant was asked by a computer screen to respond by pressing a key, to indicate whether the stimuli within the pair were the same or different in pitch contour. Only the 80% of trials in

which no key-press response was elicited were included in the analyses to avoid interference from motor-related activation.

2.3. Data analysis

The raw data were downsampled from 1000 to 500 Hz using a decimation filter, and low-pass filtered (30Hz filter with 6 db/octave) using Compumedics Scan4 Edit software. Epochs were made from -100 to 1000 ms time-locked to the beginning of the second word in each pair. Bad channels were marked at this point according to observations made during the online recording and excluded from any further processing steps. The baseline was corrected using the 100ms pre-stimulus interval and trials including artefacts of greater than +/- 75 µV from this baseline in any channel were rejected. Data in which fewer than 70% of trials remained after artefact rejection were also excluded at this point. 10 of 21 participants' data were accepted after artefact rejection.

The averaged data for baseline AA were subtracted from the averaged data for each of the deviant conditions (i.e. AB, AC, and AD)

3. RESULTS

T scores were computed between each grand average deviant condition (AB, AC and AD) and the baseline condition (AA) using Compumedics Scan4.3, and between the grand average of all deviant conditions (AB+AC+AD) and the baseline (AA).

A peak detection function was used to mark minima and maxima in a plot of T scores over time (ms), and the data were subsequently sorted and p values calculated in Microsoft Excel (peak detection was performed at 50, 100 and 250ms intervals). The results are given in Tables 1 to 3 (only results that reached significance at 95% are included; results of 100ms interval detection shown here). There were no significant results for the AB condition.

Table 1: All deviant conditions: T scores between grand averages of all deviant conditions and the baseline condition.

Channel	Latency (ms)	T score	P value
T8	56	2.99	0.015
FP2	56	2.60	0.029
O1	96	2.50	0.034
F8	58	2.45	0.037
TP8	58	2.33	0.045

Table 2: AC condition: T scores between grand averages of the deviant condition and the baseline condition.

Channel	Latency (ms)	T score	P value
T8	58	3.11	0.013
F8	58	2.62	0.028
O1	96	2.61	0.028
F4	58	2.48	0.035
FP2	62	2.44	0.038
O1	100	2.40	0.040
TP8	56	2.40	0.040
FT8	62	2.27	0.049

Table 3: AD condition: T scores between grand averages of the deviant condition and the baseline condition.

Channel	Latency (ms)	T score	P value
FP2	58	3.33	0.009
FP1	52	2.84	0.019
F8	58	2.70	0.025
FC4	946	2.40	0.040
T8	58	2.38	0.041

Scan4.3 was used to perform peak detection at 50ms and 100ms intervals on the grand averages of difference waveforms that resulted from subtracting the baseline from deviant conditions. The mean amplitude and modal latency was calculated for each interval in order to help localise peaks.

Table 4: ANOVA between difference files for deviant conditions (significance at 95% confidence interval in light grey; 99% in dark grey).

Peak & Channel	ANOVA F-ratio latency	Amplitude
P32 Fz	11.67	0.43
P58 F8	3.53	6.41
P350 T7	22.27	6.89
N520 FCz	2.43	4.61
P750 FCz	0.68	10.13
P800 FCz	0.52	9.75
P830 Cz	1.85	4.89
P890 FCz	11.86	0.92
P890 FC4	0.00	4.02
P950 Cz	9.46	26.90

This, combined with the visual inspection of the difference waveforms, resulted in a number of peaks being hypothesised, and subsequently, an analysis of variance was performed for these peaks, as shown in Table 4.

The results showed an early brain reaction (P50) to question intonation (AB), labelled (1) in Fig. 2, indicating high sensitivity to the question intonation of the deviants. Sensitivity to question intonation is also reflected in P300- and P600-like evoked potentials, labelled (2) and (4) respectively in Fig. 2.
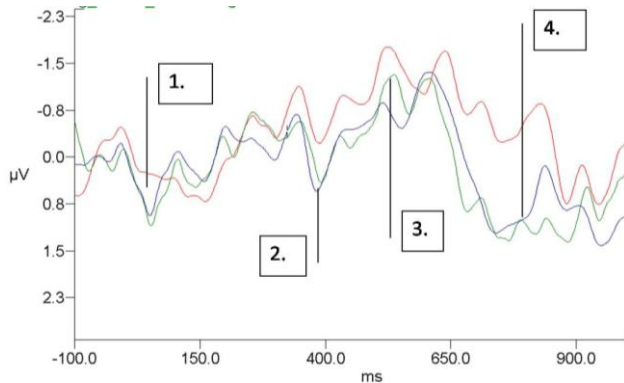

Figure 2: Comparison of smoothed grand average waveforms for the three deviants (red, green and blue) measured at electrode FC4 Comparison of smoothed grand average for difference waves for the three deviants measured at FC4, with at 1000ms top line: AB-AA; middle line: AD-AA bottom line: AC-AA.

Finally, we found an N400 for high arousal (AB) and 'neutral' question intonation (AC), labelled (3) in the figure.

4. DISCUSSION AND CONCLUSIONS

The ERP results in this study confirm that linguistic and paralinguistic processing differ both in the latency and the amplitude of activation peaks in the waveform. Categorical linguistic information in the deviant (AC and AD) elicited a set of peaks which have previously been associated with prosodic processing, e.g. [3, 12, 13, 27]. The early positivity has previously been observed with linguistic and pitch mismatch [12], while the P300-like positivity can be interpreted as a surprise reaction to the occurrence of the categorically different question intonation of the deviant [13], the P300 being associated with categorisation more generally. The P600-like evoked potential again reflects the processing of question intonation in the deviant, here of a repair function that responds when the intonation is categorically different from the baseline form in the stimuli. This is in line with the P600 being evoked in response to the processing of other categorical linguistic information in the shape of grammatical and other syntactic anomalies [21].

By contrast, gradient paralinguistic intonation only yielded significant peaks in comparison with the baseline when it was combined with a change in linguistic function (i.e. AD but not AB). The N400 which we observed in fronto-parietal areas for arousal and for question intonation (AB and AC) could be indicative of meaning difficulties in semantic integration [21].

The findings support our hypothesis that different types of intonational information recruit overlapping, but distinct neural systems, not only in terms of neural architecture, but also in the time-course of the activation of their subcomponents. That is, distinct brain areas are activated in the neural system at different points in time as different aspects of the acoustic signal are being processed in the course of abstraction from the incoming signal, becoming left-lateralised for linguistic information at later stages in processing, while paralinguistic intonation predominantly recruits right hemisphere structures.

A key implication is that categorical linguistic intonation appears to be processed very much like any other categorical linguistic information in speech in the neural system, cf. [1, 6]. This could be accounted for in a model in which speech comprehension engages two types of neurocognitive systems [7]: a predominantly left hemisphere system which is specialised to support language processing, which has evolved in addition to a distributed bilateral system which supports perceptual and cognitive processing more generally. The latter system can be seen as neurobiologically primary, while the more specialized left hemisphere system is likely to be specific to humans, according to [7].

The dual function of intonation is supported by these two distinct cognitive and neural systems; the one being encoded in the linguistic system, and the other reflecting biological imperatives much more directly. This would be in accordance with Gussenhoven [19] who proposes that linguistic uses of intonation are grammaticalised uses of universal form-meaning relations which are originally rooted in biology, but for which the encoding has become discrete, and language-specific, and for which the form-meaning relation may also have become arbitrary. Linguistic intonation is distinct from paralinguistic intonation which is governed more directly by biological imperatives.

5. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Anonymous. In press. Categories and gradience in intonation. An fMRI study. In: Romero, J., Riera, M. (eds), *The Phonetics / Phonology Interface: Sounds, representations, methodologies*. Amsterdam: John Benjamins.

[2] Begleiter, H., Platz, A. 1969. Cortical evoked potentials to semantic stimuli. *Psychophys* 6, 91-100.

[3] Boecker, K.B.E., Bastiaansen, M.C.M., Vroomen, J., Brunia, C.H.M, Degelder, B. 1999. An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology* 36, 706-720.

[4] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer; http://www.praat.org/

[5] Bolinger, D. 1970. Relative height. In: Léon, P., Faure, G., Rigault, A. (eds), *Prosodic Feature Analysis*. Paris: Didier, 109-125.

[6] Borràs Comes, J., Costa Faidella, J., Prieto, P., Escera, C. 2012. Specific neural traces for intonational discourse categories as revealed by human-evoked potentials. *J Cog Neurosci* 24, 843-853.

[7] Bozic, M., Tyler, L.K., Ives, D.T., Randall, B., Marslen-Wilson, W.D. 2010. Bihemispheric foundations for human speech comprehension, *Proc Nat Acad Sci* 107, 17439-17444.

[8] Burton, M., Small, S., Blumstein, S. 2000. The role of segmentation in phonological processing: an fMRI investigation. *J Cog Neurosci* 12, 679–690.

[9] Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

[10] Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D. 2006. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* 30, 580-587.

[11] Eulitz, C., Lahiri, A. 2005. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J Cog Neurosci* 1, 577–583.

[12] Friedrich, CK., Alter, K., Kotz, SA. 2001. An electrophysiological response to different pitch contours in words. *Neuroreport* 12: 3189-3191.

[13] Friedrich, CK., Kotz, SA., Friederici, AD., Alter, K. 2004. Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cog Brain Res* 20, 300-308.

[14] Grabe, E., Post, B., Nolan, F., Farrar, K. 2000. Pitch accent realisation in four varieties of British English. *J Phon* 27, 161–185.

[15] Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., Lurito, J. 2003a. Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain and Language* 84, 318–336.

[16] Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., Li, X. 2003b. A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Hum Brain Mapp* 18, 149–157.

[17] Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M. 2004. Hemispheric roles in the perception of speech prosody. *NeuroImage* 23, 344–357.

[18] Grandjean, D., Scherer, K.R. 2006. Examining the neural mechanisms involved in the affective and pragmatic coding of prosody. *Proc. Speech Prosody 2006*, Dresden.

[19] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.

[20] Johnsrude, I., Penhune, V.B., Zatorre, R.J. 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155-163.

[21] Kutas, M, Federmeier, KD, 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann Rev Psychol*.62: 621-47.

[22] Ladd, DR., Morton, R. 1997. The perception of intonational emphasis: continuous or categorical? *J Phon* 25, 313–342.

[23] Mayer, J., Wildgruber, D., Riecker, A., Dogil, G., Ackermann, H., Grodd, W. 2002. Prosody production and perception: converging evidence from fMRI studies. *Proc. Speech Prosody 2002*, 487–490.

[24] Meyer, M., Alter, K., Friederici, AD., Lohmann, G., von Cramon, DY. 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17: 73--88.

[25] Mitchell, RL., Elliott, R., Barry, M., Cruttenden, A., Woodruff, PW. 2003. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*. 41: 1410-1421.

[26] Obleser, J., Eulitz, C., Lahiri, A. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J Cog Neurosci* 16, 31–39.

[27] Paulmann, S., Bleicher, M., Kotz, S.A. 2013. Valence, Arousal and Task Effects in emotional prosody processing. *Front Psychol*. 4: 345.

[28] Pierrehumbert, J. 1980. *The Phonetics and Phonology of English Intonation*. Ph.D. thesis, MIT.

[29] Robin, D., Tranel, D., Damasio, H. 1990. Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language* 39, 539-555.

[30] Shriberg, EE., Ladd, D. R., Terken, J. M. B., Stolcke, A. 1996. Modeling pitch range variation within and across speakers: predicting f0 targets when 'speaking up'. *Proc ICSLP* Philadelphia, PA.

[31] Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., Ackermann, H. 2005. Identification of emotional intonation evaluated by fMRI. *Neuroimage* 24:123.

[32] Witteman, J. IJzendoorn, van, M.H. Velde, van de, D., Heuven, van, V.J.J.P., Schiller, N.O. 2011. The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsych*, 49, 3722 – 3738.

[33] Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., Li, P. 2011. Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS One* 6, e20963.