How Does The Brain Respond To The Melody Of Speech?

Using ERPs to investigate the contributions of acoustic and phonological information to prosody

Zee Yong Ng[a], Catriona Cumming[b], Brechtje Post[b], Kai Alter[a]
[a]Newcastle University, Newcastle upon Tyne, UK,
[b]University of Cambridge, Cambridge, UK

Introduction

Prosody describes the pitch contour, stress and rhythm of spoken material and has been found to contribute to both emotional and semantic meaning of language. Several studies have been carried out to investigate the effects of varying prosody on meaning and the manner in which prosodic information is processed by the human brain. Previous studies have also indicated that certain event-related potentials (ERPs) in the human electroencephalogram (EEG) correspond to the processing and reprocessing of semantic and emotional meanings of words. Although pitch contour is understood from previous work to contribute to prosody word meaning, it is yet to be found how and whether different linguistic elements of pitch contour contribute to semantics and emotional meaning in different ways. The current study used ERPs to try to elucidate the manner in which phonological and phonetic elements of pitch contour are each processed and each contribute to semantics and emotional meaning in speech.

Hypotheses

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

2.This distinction is reflected in the neurobiological processing of intonational information: They recruit overlapping, but distinct neural networks, not only in terms of neural architecture, but also in terms of the time-course of activation of the networks in question.

Method

Participants

21 Right-handed, native English participants between the ages of 17 and 32 were recruited. 11 participants' results were discarded due to excessive artefacts on the EEG recording.

EEG Recording

EEG was recorded while participants performed a categorical perception discrimination task, listening to stimuli through headphones. The EEG electrodes are placed according to the standard 10-20 system, with two mastoid electrodes. The left mastoid was used as an online reference. A SynAmps2 amplifier and Compumedics Scan 4 Acquire software were used to record the EEG. Impedance was kept below 5kΩ for the majority of EEG recordings and below 10kΩ for all recordings.

Stimuli and Task

The stimuli used were all single word utterances of trisyllabic country names in English spoken by a male native English speaker. The stimuli were resynthesised from a set of 22 base stimuli (all trisyllabic; 8 with initial stress,14 penultimate stress). For each stimulus, 4 resynthesised versions were generated, varying F0 in 4 steps: (A) fall of 3 ST, and (B) fall of 9 ST (C) rise of 3 ST (D) rise of 9 ST. Stimuli are between 1 and 1.2 seconds long. By altering the base stimuli, a 2x2 crossed paradigm was produced. The pitch contours of the stimuli are represented in Figure 1.:

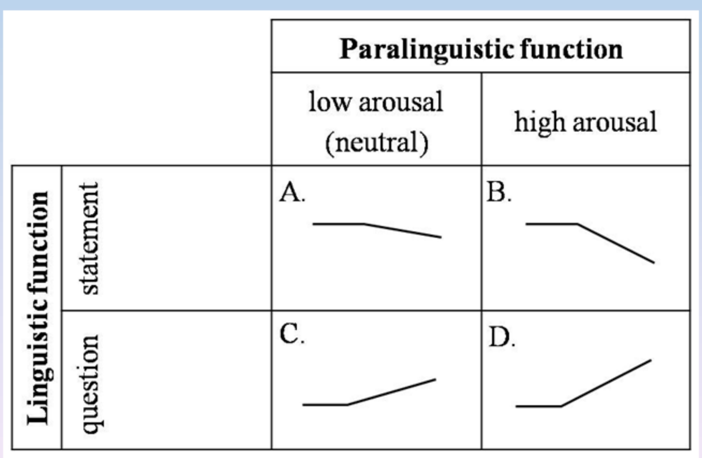
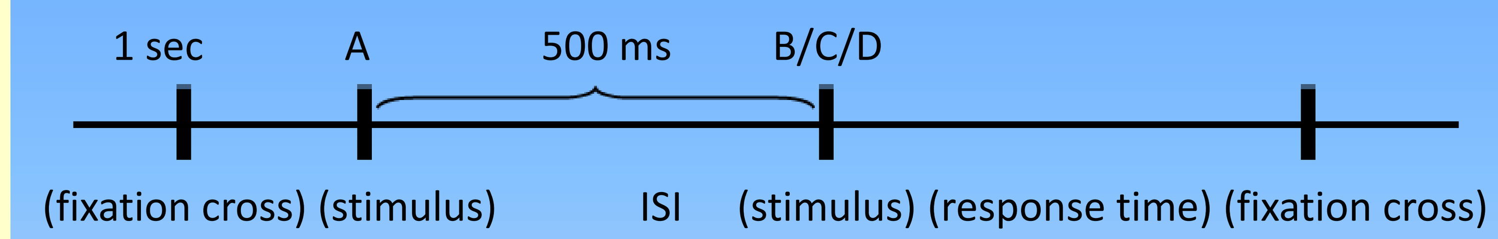


Figure 1: Representations of the pitch contours.
The baseline stimulus A had a small, downward pitch contour. This was altered in magnitude alone - changing its acoustics - to produce stimulus B; in direction alone, changing its phonology - to produce stimulus C, and in both acoustics and phonetics to produce stimulus D. Phonetic changes were proposed to inform paralinguistic function - here, arousal - while phonological changes were proposed to inform linguistic function.

Experimental Paradigm

A Categorical Perception discrimination paradigm was used to elicit 'same/different' responses. Each of the three 'comparison' stimuli (B, C, and D) were paired with the baseline (A), and A was paired with itself for the 'different' trials. Stimuli were presented in random order as follows:



For 20% of trials participants were asked to record their response by pressing a button on a button box; these were discarded from final analysis. The remaining 80% were the experimental trials which only involved passive listening.

Data in which fewer than 70% of trials remained after artefact rejection were rejected at this point and not included in the grand average. 10 of 21 participants' data were accepted after artefact rejection. Averages were made for each participant of trials using the same condition (AA; AB; AC; AD). AA was the baseline condition and AB, AC and AD the three deviant conditions. The averaged data for the baseline was subtracted from the averaged data for each deviant condition. Grand averages were computed for each subtracted file (AB-AA, AC-AA and AD-AA) and for each averaged data file (AA, AB, AC and AD).

Results

Figures 2 and 3 show the comparison of grand average waveforms generated from the data by Compumedics Scan4.3.

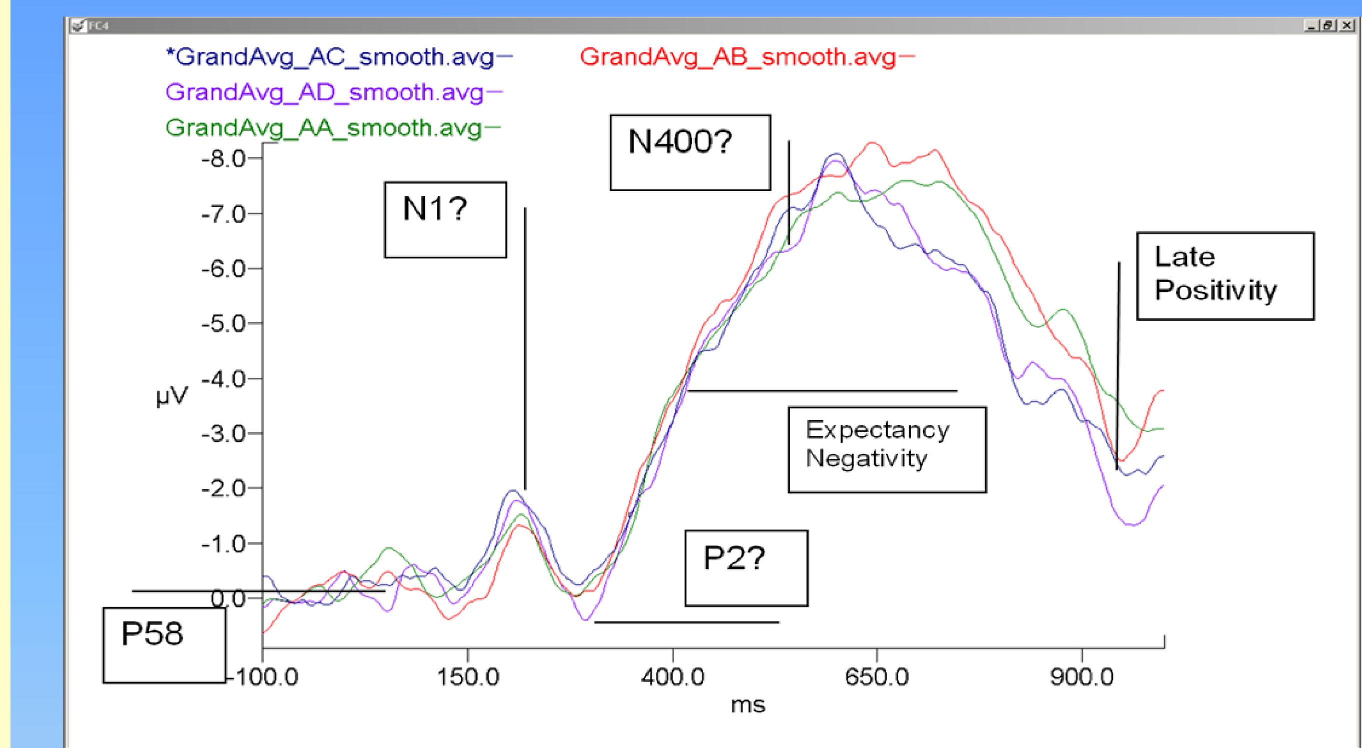


Figure 2: Comparison of smoothed Grand Average waveforms at FC4.
GrandAvg_AA (green) is the baseline stimulus; all others are deviant conditions. The N1-P2 complex is not present where it might be expected, but might be delayed with N1 peaking around 200ms and P2 around 300ms. A large expectancy negativity is seen in all files. A greater negativity is seen around 520ms in the deviant conditions compared to the baseline. Late positivity shows greater amplitude in all deviant conditions, but most pronounced in AD.

Table 1: ANOVA between difference files for deviant conditions

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

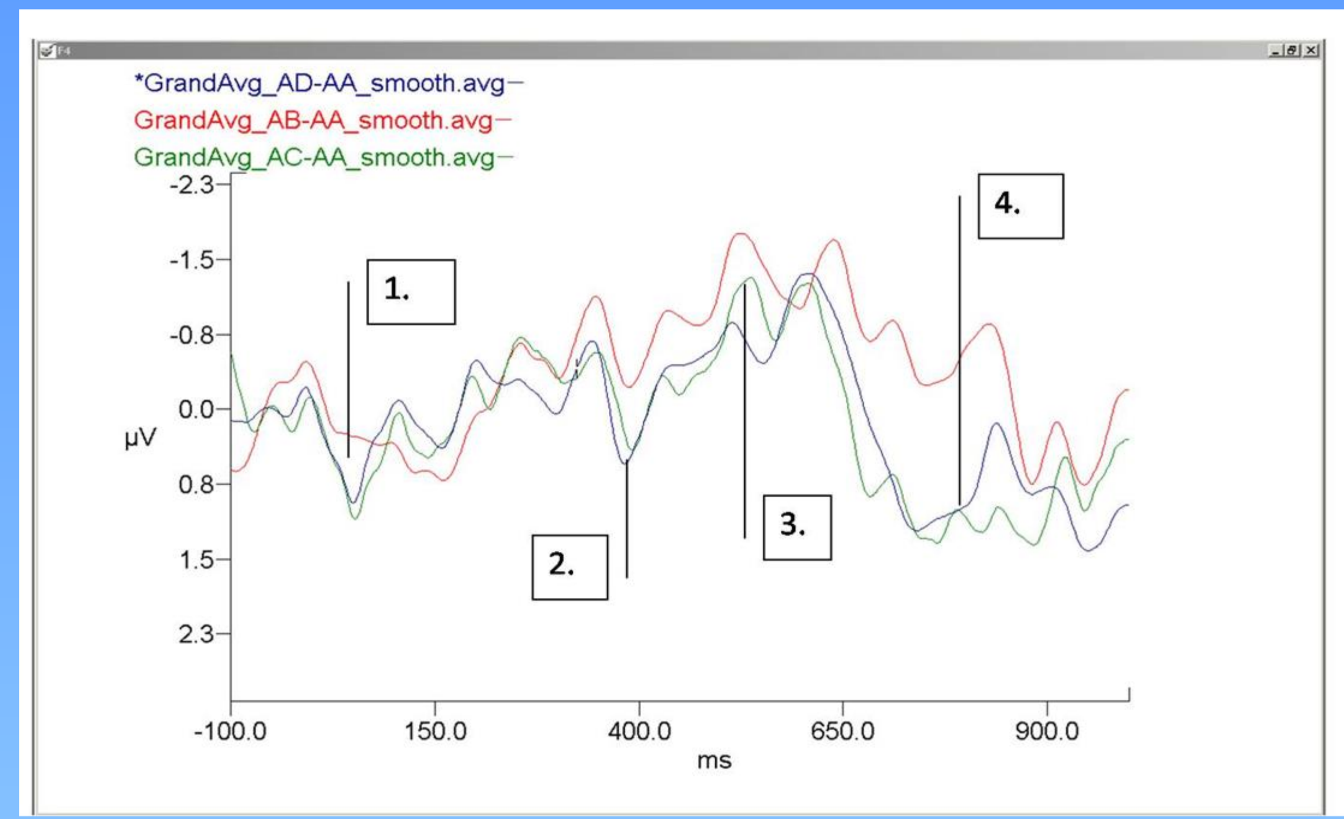


Figure 3: Comparison of smoothed grand average difference waveforms for deviant conditions – baseline.
1. Larger P58 peak seen in AD and AC conditions than AB
2. Larger P350 peak seen in AD and AC conditions than AB
3. Possible N400 peak larger in AB and AC than AD
4. Late positive potentials larger in AD and AC than AB.

Analysis of Variance

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. This combined with the visual inspection of the difference waveforms resulted in the following peaks being hypothesised:

- Positivity around 32ms in right fronto-parietal distribution (seen in AB and AD)
- Positivity between 56 and 62ms in right fronto-parietal distribution (AC and AD)
- Negativity around 320ms whose distribution differed between conditions. This was not pursued further although might be a candidate for further investigation.
- Positivity around 370ms.
- Negativity around 520ms in fronto-central distribution
- Some ambiguity around 600ms. This was not pursued although might be a candidate for further investigation.
- A number of closely positioned but distinct late positive potentials in the fronto-central distribution: P740, P800, P830, P890 and P950.

Analysis of variance between the three deviant conditions was performed for the P32, P58, P350, N520 and late positive potentials. The results of this are summarised in Table 1.

Conclusion

Phonological and phonetic representations contributes differently to meaning of language. The former contributes to linguistic meaning (semantics) and the latter contributes to paralinguistic meaning (emotion). Distinct neural networks exist for both representations as proven by late positivity (AD condition shows the greatest deflection in amplitude). Also, phonetic effects related to arousal seem to be processed earlier than phonological, categorial information.