Perception of sentence intonation in dialogues and single sentences: Behavioural and ERP measurements

Kai Alter
Newcastle University
Newcastle Auditory Group

Introduction

In intonational languages like German and English, the variation of fundamental frequency (F0) is one of the most important prosodic parameters in speech perception. This parameter is used for highlighting focussed information by pitch accentuation and for signalling phrase boundaries. Previous studies have shown that appropriate accentuation leads to efficient understanding (Bock & Mazella, 1983; Terken & Nooteboom, 1987). Studies examining the role of fundamental frequency showed moderate effects for F0-flattened speech (Laures & Weismer, 1999, Wingfield et al., 1984). But nevertheless, behavioural data do not allow a real time insight in the processing of distinct prosodic variation in the speech signal.

For a better understanding of how listeners use prosodic features, event-related brain potentials (ERP) can be measured. A specific ERP component of speech segmentation was found by Steinhauer et al. (1999). The *Closure Positive Shift (CPS)* is being evoked by IPh-boundaries which can reflect syntactic phrasing.

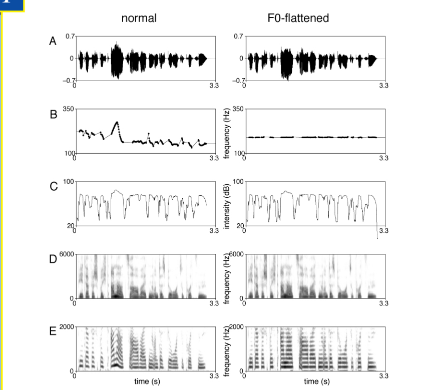
In addition to that, this component was also detectable on dialogue level after perceiving new and important information in an already given context (Hruska et al., 2001).

Two experiments were conducted to examine the role of F0-modulation on single sentence level (following the experimental setting of Steinhauer et al., 1999; see Experiment 1) and on dialogue level (following Hruska et al., 2001; see Experiment 2). In both experiments, normally intonated speech and sentences with a flattened pitch contour were presented to listeners.

Methods

Acoustic: The sentence material was read aloud by a female native German speaker in an unfocussed version or with preceding context question establishing two different focus conditions (broad and narrow). Each condition included 48 sentences with an identical word order. The material was recorded in an acoustically shielded room and was digitised with 44.1 kHz and 16 bit sampling rate. Each sentence was normalised in loudness. Furthermore, the fundamental frequency contour of all sentences was flattened to 190 Hz using PRAAT (Boersma & Weenink, 2001).

1



The pitch manipulation procedure leads to flat pitch over the whole sentences (B). The other primary prosodic features (duration and intensity) are not affected (A= oscillogram; C=amplitude envelope). The broad-band spectrogram (D) does not show any effects of acoustic manipulation, whereas the narrow-band spectrogram (E) reflects differences in the harmonic structure.

Acoustic measurements: The acoustic analysis of the speech data reflects the positions of phrase boundaries in the unfocussed conditions. In both conditions, an IPh-boundary appears after the second verb. Only in the transitive condition an additional IPh-boundary after the first verb is measurable. Each IPh-boundary is consistently characterized by prefinal lengthening, a pause insertion and a *boundary tone*. The answer sentences of the dialogue sequences show *pitch accents* on the focussed position. Furthermore, in the narrow focus condition (noun) the given information is deaccentuated. Durational marking is only used inconsistently for highlighting focussed information.

Task: Listeners had to judge whether prosody of each (answer-) sentence was appropriate regardless of global acoustic manipulation.

Psychophysiology: The electroencephalogramm (EEG) was continuously recorded from 21 cap-mounted Ag-AgCl-electrodes. The left mastoid served as reference and data where offline mathematically rereferenced against linked mastoids. The ground electrode was placed on the sternum and the electrooculogramm (EOG) were measured bipolar for vertical and horizontal eye movements. EEG and EOG recordings were amplified with a TMS-amplifier and digitised using a sampling frequency of 250 Hz. The electrode impedance were kept below 5 kΩ. Only artefact free trials were averaged per condition (3500 ms sentence presentation with a 200 ms pre-sentence baseline). The electrophysiological data were processed using the ERP evaluation package EEP (version 3.2).

Statistic analyses were primarily conducted for succeeding time windows of 300 ms over midline electrodes (FZ, CZ, PZ) in addition to regions of interest (anterior, central, posterior each for left and right hemisphere) for further topographic analyses. Moreover, baseline to peak analyses were conducted for the amplitude and latency of the N1 and P2.

Results

Experiment 1 (single sentences)

Material/Design:

The experimental items were sentences with identical word order, but conditions differ in syntactic structure (intransitive vs. transitive). In addition, items in the transitive condition were presented without a pause after the first phrase boundary. All items were presented with normal intonation and with flattened pitch.

After participants' responses to the visually presented task question, a pause of 2500 ms was inserted between two subsequent trials. In order to avoid eye movements a fixation cross was presented on the monitor starting 1000 ms before the beginning of the next trial.

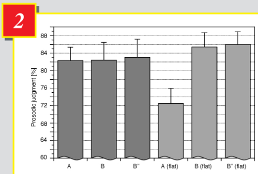
Examples:

intransitive	Peter verspricht Anna zu arbeiten $_{\text{IPh}}$ und das Büro zu putzen. *(Peter promises to work and the office to clean.)*
transitive	Peter verspricht $_{\text{IPh}}$ Anna zu entlasten $_{\text{IPh}}$ und das Büro zu putzen. *(Peter promises Anna to support and the office to clean.)*
transitive without pause	Peter verspricht $_{\text{IPh?}}$ Anna zu entlasten $_{\text{IPh}}$ und das Büro zu putzen.

Subjects:

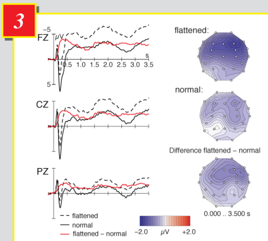
The experiment was conducted with 24 native German speaker (12 female, mean age 23, range 20-30 years). All were right handed (mean=90.8 of the Edinburgh handedness inventory; Oldfield, 1971). None of the subjects had neurological, psychiatric or hearing impairments. The median of verbal memory span amounts 35.

2



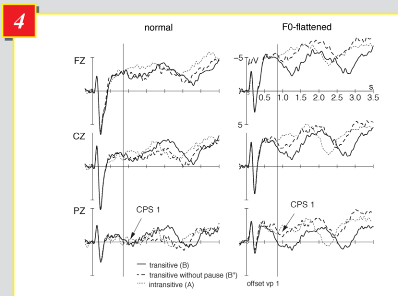
The judgment is not influenced in general by a missing fundamental frequency contour. Especially, the missing pause in B'' reveals no differences.

3



For flattened pitch items a sustained negativity at fronto-central electrodes is measurable over the whole sentence. This negativity already influences early ERP-components like N1 and P2, whose amplitudes were significantly more negative for the acoustically manipulated conditions.

4



The Closure Positive Shift (CPS) was measurable at each IPh-boundary in normal and acoustically manipulated speech with no differences. Nevertheless, if the pause is removed, the latency of the CPS is shortened. Behavioural data do not show any differences.

Experiment 2 (dialogue sequences)

Material/Design:

The dialogues (48 items per condition) were presented in a question (male speaker) - answer (female speaker) - fashion. The interval between question and answer (FAI) was 1500 ms, and the inter-trial-interval (ISI) was 2500 ms after registration of the bottom press according to the task. A fixation cross on the monitor was presented starting 1000 ms before the onset of the next trial to avoid eye movements. The examples below consist of trials as they were presented:

Examples:

Noun focus (narrow):

	Wem verspricht Peter zu arbeiten und das Büro zu putzen? *(Whom does Peter promise to work and the office to clean?)*
appropriate	Peter verspricht **ANNA** zu arbeiten und das Büro zu putzen. *(Peter promises ANNA to work and to clean the office to clean.)*
inappropriate	Peter verspricht Anna zu **ARBEITEN** und das Büro zu putzen.

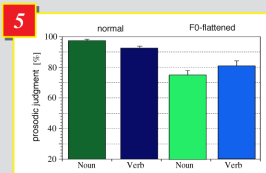
Verb focus (broad):

	Was verspricht Peter Anna zu tun? *(What does Peter promise Anna to do?)*
appropriate	Peter verspricht Anna zu A**RBEITEN** und das Büro zu putzen. *(Peter promises Anna zu WORK and the office to clean.)*
inappropriate	Peter verspricht **ANNA** zu arbeiten und das Büro zu putzen.

Subjects:

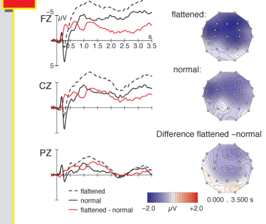
Analogous to the previous experiment, 24 native German speaking volunteers (12 female, mean age=24; range=20-30 years) took part in this study. All participants were tested on handedness (mean=91.6; SD=20.7) and verbal memory span (median=35) and none of them reported any neurological, psychiatric problems nor loss of hearing.

5



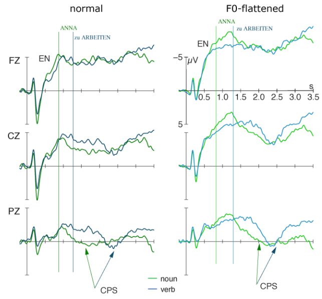
On dialogue level, flattened fundamental frequency leads to decreased identification of appropriate accentuation pattern (and inappropriate too, not described here) due to the context question.

6



On dialogue level, there is also a sustained negativity measurable in the answer sentences with pitch manipulation. Effects show statistically significant effects of PITCH in anterior and central ROIs. An additional interaction of HEMISPHERE × PITCH at central electrodes confirms a right-lateralised effect for pitch flattened conditions. Baseline-to-peak analyses of N1 and P2 show a significant decrease only for the amplitude of the P2 whereas the N1 is uninfluenced by acoustic manipulation.

7



An early frontal effect is measurable and most likely represents the expectation of the focussed information (according to Hruska et al., 2001; the so called *expectancy negativity, EN*). A positivity in dialogue sequences is measurable after perceived focussed position. This positivity is similar to the CPS on sentence level and also shows a posterior maximum. Nevertheless, in the pitch-flattened conditions the EN and the CPS after noun focus are delayed. This latency jitter might reflect the aggravated perception on dialogue level and it underlines the important role of pitch as a prosodic marker of new information.

Conclusion

In both experiments, the response to unusual prosodic features in speech is reflected by a sustained fronto-central negativity starting about 200 ms after the sentence onset (P2). This negativity has a broad distribution across the scalp and occurs in all acoustically manipulated conditions. Effects of pitch manipulation are more prominent over the right hemisphere. In comparison to other psychophysiological studies, additional brain activity for flattened pitch sentences is attributed to the missing prosodic properties of pitch and might reflect an analysis/repair-mechanism (see Meyer et al., in press).

Results show that the modulation of F0 is crucial for identifying focus accent positions in dialogues, whereas the assessment on single sentences is scarcely effected. The variation of fundamental frequency seems to be the most important prosodic cue of marking new information in dialogues. Appropriate prosodic marking leads to efficient information structural processing. However, on single sentence level a missing variation of F0 does not affect the structural analysis. Other available prosodic cues, like duration (and intensity), can also guide the listener, supporting the cue trading approach of Beach (1991).

The CPS reflects the perception of intonational phrase boundaries, being an initial marker of underlying syntactic structure, on single sentence level. Thus it might reflect the chunking of the speech input. Moreover, on dialogue level the CPS is shown only in (prosodically) appropriate dialogues and occurs after listeners have received the expected new information. For dialogues, the CPS is a marker of integration of important information. More general, the CPS on these two levels reflects the structuring of speech.

References:

Beach, C.M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644-663.
Bock, J.K. and Mazella, J.R. (1983). Intonational marking of given and new information. Some consequences for comprehension. *Memory and Cognition, 11*, 64-76.
Boersma, P. and Weenink, D. (2001). PRAAT. A system for doing phonetics by computer. *Institute of Phonetic Science*. (www.praat.org).
Hruska, C., Alter, K., Steinhauer, K. and Steube, A. (2001). Misleading dialogue: Human brain's reaction to prosodic information. *Paper presented at the Oralite et Gestualite, Aix en Provence, France.*
Laures, J.S. and Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at sentence level. *Journal of Speech, Language, and Hearing Research, 27*, 128-134.
Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D. and von Cramon, D.Y. (in press). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*.
Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychology, 9, 97-113.
Steinhauer, K., Alter, K. and Friederici, A.D.(1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience. 2 (2)*, 191-196.
Terken, J. and Nooteboom, S.G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2(3/4)*, 145-163.
Wingfield, A., Lombardi, L. and Sokol, S. (1984). Prosodic features and the intelligibillity of accelerated speech: Syntactic versus periodic Segmentation. *Journal of Speech Language and Hearing Research, 27*, 128-134.