

Jörg Peters
Radboud University Nijmegen

Rare pitch patterns

What frequency can tell us about intonational meaning

For those working on prosodic features of spontaneous speech the following will sound familiar. On the one hand, speakers use pitch patterns which are well known, frequently occurring, and highly systematic. On the other hand, speakers use pitch patterns which are less known, infrequent, and usually are regarded as an unwanted source of intra- and interspeaker variation. Those infrequent pitch patterns are often ignored or reported as marginal cases which are irrelevant to the research question.

This paper shows that rare pitch patterns can be a valuable source for both developing and evaluating compositional theories of intonational meaning. In free conversational speech, rare pitch patterns can be expected to occur in more restricted contexts than more frequent patterns. In our view, any well-designed compositional theory of intonational meaning must characterize phonologically distinct pitch patterns by sets of abstract semantic features which account for their usage in more or less restricted contexts. In this way, the theory will predict relative frequency distributions of pitch patterns. We illustrate this view with the help of German conversational speech data and propose a compositional theory of intonational meaning which accounts for the frequency distribution found in our corpus. Our analysis shows that spontaneous speech is important for research on intonational meaning not only by providing the necessary contextual information to understand intonational functions. It also provides valuable frequency information, which may reflect semantic complexity.

Methodological aspects of separating systematic and individual prosodic features

Prosody has proved to be a highly variable system, influenced by stylistic, emotional and dialectal effects. Descriptions of 'the' system of a specific language, a specific style or a specific dialect are necessarily reductions of this variability. However, idiosyncratic peculiarities of the individual speakers can – in the best case – blur the systematic variation or they can – in the worst case – bias and distort the description. On the other hand, these characteristic features allow for an identification of the individual. This singling out of individuals is used for forensic purposes where the methods on the segmental level are developed and continually improved. As the linguistic description of the prosody is not yet as refined as that of the segmental level, this identification cannot yet be very accurate. A description of the prosody that takes into account individual features right from the beginning can alter both aims: the linguistic description and the forensic recognition. However, the methodological aspects have hardly been addressed.

The paper will present a method of clustering speakers and singling out individuals on the basis of prosodic features. The discussion is based on a database for a description of Swiss-German prosody. Accordingly, the main goal of this clustering can be viewed within a variationist approach. It is a) to give an extraction of data that best represent the prosodic features of a variety, b) to document the variation possible within that variety, and c) to separate related varieties. Identifying systematic and individual differing traits is therefore a must. The paper focuses on the methodological implications of this approach for research on prosody.

Corpus-based analysis of prosodic and segmental features influencing F_0 peak alignment

We present an approach specifically oriented towards determining phonetic detail in an effective manner. The speech database of the IMS German Festival synthesis system (1 male speaker, 160 minutes of speech, prosodically labeled using GToBI) serves as a corpus for an investigation of the alignment of peaks in H*L pitch accents in German (2681 instances in the corpus). The procedure allows the detection of correlations between the phonetic implementation of a tonal category and virtually all levels of the prosodic hierarchy (from phrasing to accent distribution to syllable structure) and segmental environment. Festival parameters provide the descriptive framework. Peak measurement is achieved by automatically locating the F_0 peak in a syllable labeled with a H*L pitch accent.

Despite the fact that this broad approach lacks a strictly controlled tonal, segmental or syllabic environment and is based on a corpus provided by just one speaker our results confirm established insights from earlier studies dealing with languages other than German. This concerns for example the special status of nuclear pitch accents, as peak alignment in such accents is shown to occur significantly earlier than in non-nuclear accents (mean: 38% vs. 53% of syllable duration) as well as the effect of tonal repulsion, corroborated by the fact that a nuclear pitch accent in the last syllable of the intonation phrase is aligned significantly earlier in the syllable (mean: 21%) than a non-final nuclear pitch accent (mean: 44%).

Similarly, the influence of syllable structure investigated on the basis of the Van Santen/Hirschberg classification (sonorant, voiced obstruent, voiceless obstruent) is shown to be significant both for onsets and codas, sonorant codas leading to later peaks, while peaks are earliest for voiceless obstruents. For onsets the effect is the exact opposite (peaks occur earliest with sonorant onsets).

Trudel Meisenburg & Christoph Gabriel
University of Osnabrück & University of Hamburg

**Exploring variation and change in Romance intonation: Occitan and French in
Southern France, Spanish and Italian in Argentina**

Though all deriving from Latin, Romance languages display a fair amount of variation in their accent and intonation systems, which seems at least partly due to the more or less consistent contact between different vernaculars. Such contact situations persist since centuries between French and Occitan in Southern France, and since the late 19th century between Spanish and Italian in Argentina. While the Italian influence on Argentinean intonation is largely uncontested but still requires intensive research, a possible influence of Occitan, which could contribute to explain the highly divergent nature of French prosody, has not yet been sufficiently discussed. In a pilot study Hualde (2003) signals an important parallel between Occitan and French: In addition to metrically strong syllables the Occitan speakers he examined made use of other positions for intonational structuring. Like the phrase-initial pitch movements known from French (Jun/Fougeron 2002), these additional pitch events mainly hit the first syllable of content words. The Italian influence on Argentinean intonation, on the other hand, is characterized by the realization of pre-nuclear pitch accents as H* (instead of standard Spanish L*H, Colantoni/Gurlekian 2004) and a descending final contour usually labeled as ‘long fall’ (Kaisse 2001), both typical Italian features absent from other Spanish dialects.

Our contribution will focus on the methodological problems encountered when studying intonation in speech situations that partly result from language contact. As for Argentinean Spanish, the main concern will be to conceive the gradation within Argentinean intonation and to isolate different subgroups of speakers according their age and socio-cultural affiliation. For Occitan on the other hand, a major problem lies in the recruitment of adequate subjects: As the rare primary speakers of this highly endangered language are neither used to speak Occitan with any one they are not very well acquainted with nor are commonly familiar with the written form of their language, a special experimental design has to be developed.

Klaus J. Kohler
University of Kiel

Analysing speech rhythm in different languages - A new research paradigm

The languages of the world are frequently allocated to *stress timing* or *syllable timing* rhythm. Phoneticians have collected acoustic data from text readings and measured feet or syllable durations, which they have assumed to form isochronous patterns. But the degree of variability in articulatory sequencing, tolerated for maintaining rhythmicity, or responsible for destroying it, cannot be established in this way. Rhythmic grouping needs perceptual testing by systematic variation of f_0 , duration and acoustic energy in elementary syllable chains. This is the first part of the new paradigm.

This approach follows from looking at rhythm as a feature of speech communication, which a priori includes the listener. It has a guiding function: the temporal recurrence of similar signal stretches helps listeners to decode speech more efficiently because incoming signals can be projected onto an ongoing pattern. Rhythm is a dynamic bodily action that transmits varying degrees of regularity in pitch, time and intensity for a listener, not isochrony of static units (mora, syllable, foot). Rhythm is phonetic performance operating on prosodic structures set by the language. They may be highlighted or overruled in performance; there are good rhythmical speakers and bad ones.

In the second part, data from trained and untrained speakers reading a standard text provide an intuitive spread along a scale from good to bad rhythmicity, and are experimentally evaluated by listeners. The resulting clustering of the data is analysed in terms of syllabic timing, pitch and energy patterns, and related to the results from the perceptual experiments in the first part of the paradigm, leading to hypotheses about margins of regularity for rhythmic speech. They enter systematic parameter manipulation in speech synthesis for perceptual assessment of the degradation or improvement of examples originally classified as good or bad rhythmic speech, respectively.

The third part of the paradigm involves the parallel investigation in two languages that are diametrically opposed as to their intuitively perceived rhythmicity: German and French.

The paper will outline the research paradigm on the basis of some performance data.

Rebecca Hughes & Beatrice Szczepek Reed
University of Nottingham

Interpretation of prosodic cues in read-aloud vs spontaneous speech data

The gating paradigm has been used by Grosjean (1983) and Grosjean and Hirt (1996) to investigate whether listeners who hear increasing lengths of a read-aloud sentence can predict whether that sentence is complete after reaching its first syntactic completion point, or whether it will continue, and if so, for how many words. Their findings suggest that subjects accurately predict the length of a read-aloud sentence-in-progress, and that those predictions are based on prosodic cues within the potentially last word.

This paper reports on an attempt to recreate Grosjean's experiments with recordings of spontaneous, rather than pre-scripted data. Three data sets were used: spontaneous dinner conversations, radio phone-in programmes and read-aloud sentences. None of the experiments could replicate Grosjean's findings. Instead, our results illuminate the differences between laboratory data and those recorded from real-life conversations. The very process of adapting both the methodological setting and the data in order to achieve a meaningful comparison raised questions concerning comparability of results gained from different data sets. The findings also raised issues surrounding the interpretation of intonational cues in real time: while the laboratory setting manages to stabilize all parameters and focus on intonation alone, the nature of naturally-occurring language as locally emerging and context dependent means that the status of utterances as complete or incomplete is under constant interactional negotiation, and intonation works alongside a cluster of cues which signal potential completion.

Grosjean, F. (1983) 'How long is the sentence? Prediction and prosody in the on-line processing of language' *Linguistics*, 21, 501-29.

Grosjean, F. and Hirt, C. (1996) 'Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects, *Language and Cognitive Processes*, 11, 107-134.

Kai Alter
Newcastle University

**Differences between single sentence and dialogue processing:
Evidence from Evoked Potentials**

Speech contains prosodic cues that are associated with linguistic (e.g. lexical, semantic, syntactic and phonological) information. Acoustic analyses of a multitude of speech stimuli in speech production studies have shown that these cues consist of systematic variation in pitch, duration, and loudness. Surprisingly little research has addressed the question of how the human brain uses prosodic information in sentential processing. We investigated two aspects of this question:

- 1st: The brain's reaction to phrase markers in speech and music, and
- 2nd: The brain's reaction to accented/prominent constituents.

Electrophysiological data were used to explore the brain basis of the processing of prosodic information. Initial studies on the processing of sentences (Steinhauer, Alter & Friederici 1999) have revealed a positive-going shift in the Event Related Brain Potentials at phrase boundary positions – the Closure Positive Shift (CPS). These studies demonstrate the strong influence of prosodic information on syntactic parsing. Further studies examined the processing of boundary markers dependent on syntactic and segmental information contained in the speech signals. In these studies, naturally spoken sentences were compared to meaningless jabberwocky sentences containing morpho-syntactic information, and hummed sentences only preserving the original intonation contour. The CPS was observed at phrase boundaries for meaningful sentences as well as for meaningless (jabberwocky) and hummed sentences.

We also investigated the effects of accentuation and prominence on the processing of incoming speech material in dialogues to gain insight into the brain reactions to the processing of accents/prominence. We observed a CPS-like component at positions in sentences that are highlighted by means of prominence.

The data indicate that prosody has an important supportive in auditory sentence processing.

Steinhauer, K., Alter, K. & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191-196.

Intonation in contrast: spontaneous speech elicited by 4 different techniques

Prosody varies according to discourse setting, therefore one needs to be in elicitation techniques selection. The goal of our Swiss National Science Foundation Project is to study the prosody of 4 Swiss German dialects. We used 3 different techniques of eliciting spontaneous speech. During the recording sessions, the subjects, gymnasium students, performed a spontaneous interview with the researcher, a game description, and the description of a picture.

In the interview, the speaker is able to realize comparatively long passages without interruption. The disadvantage is the varying emotional content of the elicited speech, which in turn affects pitch representation. The game description task is sensible in that lengthy passages of spontaneous speech are elicited – the drawback being the interviewee's error awareness of explaining the game, which, again, affects pitch. As for the final elicitation technique, there is one major downside: speech produced by a picture description is often list-like in character, i.e. the acoustic correlate of F0 being a final rise in declarative sentences.

It is evident how each of these techniques bears advantages and disadvantages. The question rises whether data elicited by the spontaneous interview, for example, can be compared with data elicited by the picture description. By means of the Fujisaki model (Fujisaki 1988), the typically resulting intonation contours of the 4 elicitation techniques will be contrasted. A direct link will be established as to the repercussions of these findings on our project and how we tackled the problem of intonational variation due to differing elicitation techniques.

Fujisaki, Hiroya. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In Fujimura, O. (ed.), *Vocal physiology: voice production, mechanisms and functions*. New York: Raven. 347-355.

Sandra Döring
University of Leipzig

On Prosodic Aspects of Parenthetical Constructions in Modern German

This paper will focus on prosodic aspects of parenthetical constructions in modern German, particularly on the transition zone between the first part of the anchor clause and the parenthetical construction, and the related question of intonational phrasing.

In the phonetic-prosodic part of the investigation, the main focus will be on how the transition from the first part of the anchor clause to the parenthetical construction is prosodically realised. The phonological-prosodic research interest, however, will lie in the problem of prosodic domains: Which prosodic domain precedes the parenthetical construction, and which prosodic domain follows the construction? Which prosodic domains constitute parenthetical constructions?

The analysis will be restricted to not-integrated parenthetical constructions (non-integratable and integratable phrases) and will be based on spoken examples taken from the debates held in the German House of Parliament.

References

- Döring, Sandra (2007). "Quieter, faster, lower, and set off by pauses? Reflections on prosodic aspects of parenthetical constructions in modern German". In Dehé, N. & Y. Kavalova (eds). *Parentheticals*. Amsterdam. Benjamins.
- Nespor, Marina & Irene Vogel (1986). *Prosodic Phonology*. Dordrecht. Foris.
- Selkirk, Elizabeth O. (1984). *Phonology and Syntax*. MIT.