

Prosodic matching of response tokens in conversational speech

Jan Gorisch

University of Sheffield, Department of Computer Science, Department of Human Communication Sciences

The phonetic forms and discourse functions of response tokens such as *mm* have been investigated for several decades. In general, previous studies have attempted to classify response tokens according to their communicative function, and then to search for prosodic similarities within each class (Gardner, 2001). Using data from naturally-occurring research meetings in the AMI corpus (native British and American speakers) we investigated the response token *uhu* in this way. Interactional analysis was used to identify instances of *uhu* that functioned either as continuers or as acknowledgments, and then we attempted to identify prosodic (and visual) cues which might differentiate these two functions. No difference in pitch patterns (F0 range, F0 movement) could be found between response tokens with the two different conversational functions: pitch characteristics could vary between extremes (e.g. rising and falling pitch movements) within classes that were distinct from an interactional point of view. Visual cues such as gestures and gaze did not predict these differences either. However, it was observed that pitch characteristics of tokens of the same class appeared to depend on the pitch characteristics of the immediate prior talk of the interactional partner. If, for example, the prior talk ended with rising intonation, the response token was also produced with a rise if it was encouraging the other speaker to continue talking. In order to perform the same action, the intonation of the “*uhu*” was falling if the previous talk ended with a fall. This was not the case when the utterer of the “*uhu*” was taking the floor or projecting to do so. In order to quantify these effects, a technique that objectively measures the similarity of prosodic features such as pitch movement and individual speakers' ranges is presented. Using some examples we explore how copying vs. non-copying behaviour in the prosodic domain of short response tokens is used to manage talk in multi-party interaction (Szczepek Reed, 2006). The analysis is based on the principle of cross-correlation where the similarity of two signals – here F0 contours – is established. The method is extensible to cover other prosodic cues such as intensity or tempo, and even visual cues such as head nodding.