

# **Inferring meaning of listener vocalizations through a perception study**

**Sathish Pammi and Marc Schröder**  
**DFKI GmbH, Germany**

The vocalizations of listeners in a dialogue are an important but little studied aspect of social interaction. Such vocalizations include pure backchannels such as *uh-huh* or *m-hm*, as well as feedback signals providing additional information about the listener's understanding, affective state, social stance, etc.

In this study, we investigate the vocalizations of four British English listeners in a database of spontaneous dialogue speech. A total number of 441 listener vocalizations were extracted from the conversation. Each vocalization was annotated with a single-word description such as '*myeah*' or '*laughter*', and an intonation contour was automatically computed by fitting a 3rd-order polynomial to f0 values extracted using an autocorrelation-based pitch tracker. Separately for each speaker, we used K-means clustering of intonation contours to identify the vocalizations with a similar prosody.

In order to investigate the respective perceptual effect of prosody and segmental form, two sets of stimuli are extracted from the clustered data. On the one hand, stimuli with the same prosody and varying in segmental form (as determined from the single-word description); on the other hand, stimuli with the same segmental form but varying in prosody. The stimuli are presented in a web-based listening experiment. Participants can choose among a range of meaning categories such as: *agree, understand, amused, friendly, interested, etc.*

By the time of the workshop, we will have conducted the listening experiment and will report about the findings.