

# Die Erkennung von konversationellen "Hot Spots" auf Grundlage von akustischen Analysen und Bewegungsdaten

Catharine Oertel, Petra Wagner  
[catharine.oertel@uni-bielefeld.de](mailto:catharine.oertel@uni-bielefeld.de), [petra.wagner@uni-bielefeld.de](mailto:petra.wagner@uni-bielefeld.de)

Bei der Betrachtung natürlichsprachlicher Konversationen fällt auf, dass Dialogpartner ihren sprachlich-prosodischen Ausdruck häufig aufeinander abstimmen, z.B. hinsichtlich der Sprechrate [1]. Ferner sind natürliche Gespräche durch eine Dynamik geprägt, also durch ein phasenweise mehr oder weniger starkes Engagement der Gesprächsteilnehmer. Momente, in denen die Gesprächspartner besonders rege an der Konversation teilnehmen, werden als *Conversational Hot Spots* bezeichnet [2]. Das Abstimmungsverhalten zwischen Gesprächspartnern hinsichtlich ihres prosodischen Ausdrucks legt nahe, dass auch solche Hot Spots prosodisch gekennzeichnet sein sollten. Eine automatische Erkennung solcher Hot Spots kann sinnvoll sein, z.B. zur Identifikation von Teilnehmerrollen im Dialog, zur Konfliktresolution in Besprechungen oder auch zur Beschleunigung der Suche in multimodalen Korpora. Wir verfolgen die Erkennung von Hot Spots auf der Basis nicht funktionaler, also rein akustischer Signaleigenschaften sowie Bewegungsdaten, da die Automatisierbarkeit der Methode im Vordergrund steht. Hierfür wurde zunächst ein Korpus natürlichsprachlicher Konversation erstellt. Dieses Korpus wurde an zwei Tagen in je 4-stündigen Aufnahmesitzungen aufgenommen. Während der Aufnahme wurden eine Vielzahl akustischer, visueller und motorischer Daten erhoben. Es wurde darauf geachtet, dass das Korpus hinsichtlich des Alters, Geschlechts und des Bekanntheitsgrades zwischen den Sprechern ausbalanciert ist. Die Konversation wurde weder in irgendeiner Weise geleitet noch wurde ein aufgabenorientiertes Szenario entwickelt [3]. Zusätzlich wurde ein Annotationsschema für Hot Spots entwickelt, welches auf möglichst hörerunabhängigen, intuitiven Eindrücken basieren sollte und kein Training erfordert. Das Annotationsschema wurde perzeptiv validiert und zu den akustischen Analysen sowie Bewegungsdaten in Beziehung gesetzt. Erste Ergebnisse lassen den Schluss zu, dass es möglich ist, konversationelle Hotspots auf der Basis von mittlerer Tonhöhe (ERB), Grundfrequenzumfang, automatisch detektierter Sprechrate, RMS-Intensität und automatisch detektierter Silbenprominenz zu identifizieren.

## Literatur

- [1] Giles, H., Coupland, N., & Coupland, J..1991. Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge, England: Cambridge University Press.
- [2] Wrede, Britta & Elisabeth Shriberg. 2003. Spotting "Hot Spots" in Meetings: Human Judgments and Prosodic Cues. *Proc. Eurospeech 2003*, 2805-2808.
- [3] Oertel, Catharine, Fred Cummins, Nick Campbell, Jens Edlund & Petra Wagner. 2010. D64: a corpus of richly recorded conversational interaction. *Proc. LREC 2010 .Workshop: Multimodal Corpora–Advances in Capturing, Coding and Analyzing Multimodality. 2010.*