

The role of accommodation in expressing emotions to newborn babies

Katalin Mády, Uwe D. Reichel, Anna Kohári & Ádám Szalontai

Hungarian Research Centre for Linguistics, Budapest

Introduction and aims: An important function of infant-directed speech (IDS) is expressing positive emotions towards the baby, manifesting in prosodic features such as lower speech rate, higher f_0 , larger pitch range etc. (see [3] for Hungarian). Longitudinal changes in these parameters were investigated in mothers when talking to their own infants [2] in comparison to adult-directed speech (ADS). The most striking change throughout this period was the difference between the first recording directly after giving birth on the one hand, and the ones at 4, 8 and 18 months of the baby, on the other. The absence of expected prosodic cues with newborn babies can be due to language-internal and language-external factors: missing accommodation to the newborn baby, or the special emotional status of the mother: stress, fatigue, hospital environment etc.

In order to dissociate the influence of these factors, we conducted a study in which mothers from the longitudinal data set, who gave birth to their first child (primipara group, PP), were compared to mothers of newborn who already had at least one child (multipara group, MP). If they use more IDS with their 0 month-old baby in the hospital, than PP mothers do, the role of accommodation (its presence in MP and absence in PP) has a stronger explanatory power on the longitudinal differences than environmental factors.

Single prosodic parameters as f_0 or energy cannot be directly related to certain emotions. For example, higher f_0 characterises both elation joy (positive) and fear (negative), while comfort (positive) and boredom (negative) are both reflected by lower f_0 . Thus, prosodic parameters need to be interpreted in a more complex manner. In this study the emotional load of IDS was investigated perceptually, by using a two-dimensional arousal-valence space [1].

Methods and materials: 22 primipara and 23 multipara mothers were recorded in the same hospital 1 or 2 days after giving birth to their baby. They told a fairytale semi-spontaneously first to the experimenter (ADS), then to their own baby (IDS) (for more details and an acoustic approach, see [3]). Four female participants labelled 6 fixed utterances from all participants both in IDS and ADS ($n = 1908$). Arousal and valence scores ranged -4 to $+4$, with 0 meaning ‘neutral’.

Results: Krippendorff’s alpha for inter-rater agreement was rather low: 0.4074 for arousal and 0.243 for valence (unlike scores, tendencies were comparable between labellers). Thus, linear mixed-effects models were set up to use register (IDS vs. ADS) and parity (PP — MP) as fixed effects and speaker, utterance and labeller as random effects (intercept-only models) in R using the lme4 and car packages. Arousal scores were significantly higher in IDS in both the PP and MP group ($p < 0.001$). Parity did not have a main effect, but its interaction with register did ($p = 0.003$), showing that the ID-AD difference was considerably larger in the MP group. Significantly different valence scores were only found for register, but not for parity or its interaction with register (see Figure 1).

Discussion: Results show that mothers express stronger emotions when talking to their infant already from birth. Multipara mothers who are already experienced in IDS communication with their other children adapt their speech behaviour to the newborn baby, while primipara mothers still develop this skill with growing age of their baby. Thus it seems that psychical and environmental factors (stress, fatigue and hospital) cannot explain the longitudinal results [2].

In perceptual rating experiments (e.g. [4]), valence was less well recognized by speech acoustics alone than arousal. This finding is generally confirmed by cross-modal comparisons of acoustic and visual features in automatic prediction of these emotion dimensions (e.g. [5]): while arousal is well manifested in the acoustic speech signal, valence is better predicted from facial expression and spoken content.

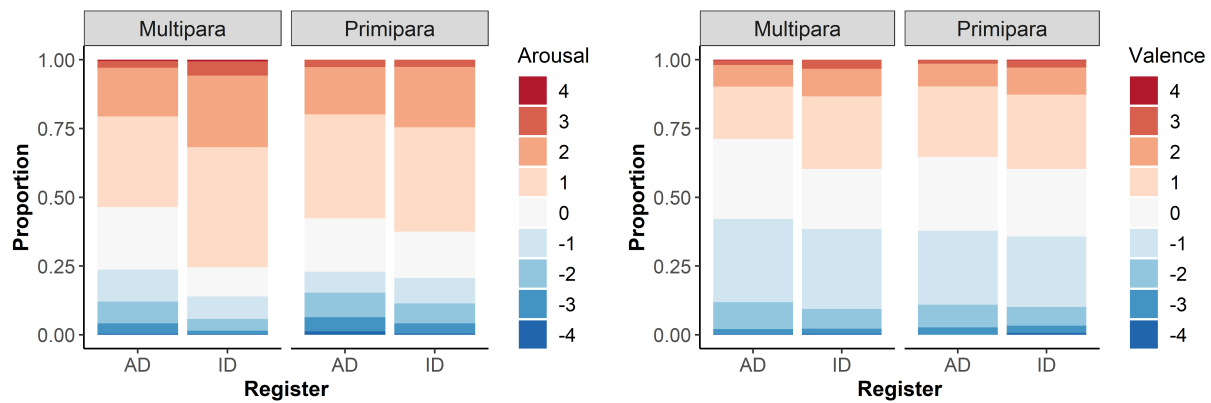


Figure 1: Scores for arousal and valence.

Acknowledgements: This study was funded by the National Research, Development and Innovation Office, grants K 115385, K 135038 and PD 134775.

References

- [1] T. Giannakopoulos, A. Pikrakis, and S. Theodoridis. A dimensional approach to emotion recognition of speech from movies. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pages 65–68, 2009.
- [2] Katalin Mády, Uwe D. Reichel, Anna Kohári, Andrea Deme, and Ádám Szalontai. Primary functions in infant-directed speech and their longitudinal development. In *Proc. Beszédkutatás – Speech Research, Budapest*, pages 62–64, Budapest, 2020.
- [3] Katalin Mády, Uwe D. Reichel, Ádám Szalontai, Anna Kohári, and Andrea Deme. Prosodic characteristics of infant-directed speech as a function of maternal parity. In *Proc. 9th Speech Prosody, Poznań*, pages 294–298, Poznań, 2018.
- [4] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The measurement of meaning*. Number 47. University of Illinois press, 1957.
- [5] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017.