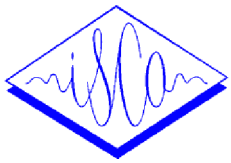

Speech Prosody 2006
Dresden, Germany
May 2-5, 2006

Comparing vocal parameters in spontaneous and posed child-directed speech.

F. Schaeffler, V. Kempe & S. Biersack

Department of Psychology
University of Stirling, Scotland, UK
{felix.schaeffler; vera.kempe; sonja.biersack}@stir.ac.uk

Abstract

Research on the facial expression of emotion distinguishes between correlates of posed vs. spontaneous emotion expression. Similar research in the vocal domain is lacking. In this study, we compare changes in a range of vocal parameters between posed vs. spontaneous adult-directed (AD) and child–directed (CD) speech. CDS is a highly affectively charged speech register which lends itself well to the study of posed vs. spontaneous emotion expression. A group of mother addressed an adult and their child, and a group of non-mothers addressed an imaginary adult and an imaginary child. The results confirm adjustments in pitch, formants and speech rate typically reported for CDS in both groups. At the same time, they show that source parameters not in service of linguistic function, such as shimmer (perturbations in fundamental period amplitude) and harmonics-to-noise ratio show clear group effects suggesting that they may constitute veridical indicators of spontaneous emotion expression.

1. Introduction

Facial and vocal expressions are the primary means of conveying emotions. Emotion expressions are not only an epiphenomenon of the speaker's emotional state, but may also serve a direct signaling purpose. It has been suggested that the expression of emotions could be a powerful means for the enforcement of communicative goals and the manipulation of the interlocutor [1].

Extensive research in the facial domain has highlighted the distinction between spontaneous and posed emotion expressions. Facial expressions of posed emotions differ from the expression of spontaneous emotions ([2], [3]), and the differences in hemiface involvement and timing have been attributed to differences in hemispheric control.

In the vocal domain, the same channel is used for linguistic communication and emotion expressions. This is an important contrast to the facial domain (in normal hearing humans). There is, however, evidence that speech and vocal emotion expression use different sensory-motor systems [4]. It is therefore quite likely that even in this domain differences between posed and spontaneous emotion expression can be observed, although spatial lateralisation effects are more restricted in vocal signals [5] than in facial expressions and linguistic and non-linguistic signals are expressed using the same anatomical structures.

Differences in the acoustic features of posed and spontaneous emotions have so far not been addressed directly. One of the reasons is that research on vocal emotion expression has mainly used acted emotions [6]. This has been justified by the assumption that emotion expression is governed by display rules, i.e. rules of permissible expressions that vary between cultures, which are acquired during development [7]. It has been argued that even spontaneous emotion expressions contain elements of deliberate emotive signaling [6], which can modulate the form and intensity of emotion expressions.

However, relying mainly on the study of acted emotions obscures the potential difference between posed, controlled, deliberate emotive signals on one hand, and involuntary, spontaneous emotion expression on the other hand. In this study, we address this issue by contrasting the acoustical correlates of deliberately produced emotion expressions with correlates of the expression of induced genuine emotions. We compare posed and spontaneous child-directed speech (CDS). CDS is a strong emotionally charged speech register. There is evidence from neuro-imaging studies that interaction with their own baby serves as a powerful inducer of positive mood in mothers [8]. Thus, CDS of a mother can be regarded as a genuine expression of positive affect towards her child.

Genuine CDS is contrasted with CDS of non-mothers directed towards an imaginary child. Previous studies by Jacobson et al. [9] and in our lab [10] have shown that CDS directed towards an imaginary child exhibits many typical characteristics of genuine CDS such as raised pitch, wider pitch range and slower speech rate, albeit in an attenuated fashion. Thus, it can be regarded as an instance of posed emotion expressions as speakers use their knowledge about the speech register to mimic genuine CDS. In the present study, we will explore the differences in vocal correlates of CDS in mothers addressing their own child with non-mothers addressing an imaginary child.

We predict that vocal parameters which, in any given language, are in service of linguistic prosody and, thus, are likely to be controlled by the left hemisphere, such as f0 and the formants, will be used for posed emotion expression, while vocal parameters not in service of prosody such as voice quality parameters (timbre, perturbations etc.) may serve as faithful indicators for genuinely experienced emotions.

2. Method and Subjects

2.1. Speech data and subjects

The material of the study was taken from a larger study on prosodic disambiguation in CDS. Two groups of speakers participated in the experiment: a group of 24 mothers with infants between 23 and 45 months of age, and a group of 24 women without children (henceforth "non-mothers"). The mother and the non-mother group were roughly matched for age and were between 23 to 46 years of age. All participants spoke the Scottish variety of British English.

The task consisted of a set of commands towards an interlocutor to touch a certain soft toy from a group of soft toys. Some of these commands were ambiguous and demanded prosodic disambiguation by the participants. Other commands were unambiguous. The test phrase analysed in the current study was taken from the unambiguous set of instructions. All test phrases and the intended actions were presented in a booklet, and all participants were encouraged to address their interlocutor and not just read the sentences aloud.

The task was performed in two conditions. In one condition (the ADS condition), the participants addressed an adult confederate. In the other condition (the CDS condition), the mothers addressed their toddlers, and the non-mothers addressed an imaginary child of similar age. The children of the mother-group were present during the entire test sequence, i.e. also during the ADS condition.

The two tasks were performed in immediate succession. To account for potential order effects, half of each group performed the task starting with the ADS condition followed by the CDS condition and the other half performed the CDS condition first.

2.2. Measurements

The analysed phrase was "*Touch the snake and the fish*". Prosodic features were extracted from the vowels of the content words "touch", "snake" and "fish". (In the Scottish variety, the vowel in "snake" is usually a monophthong [e:].) The following linguistic and prosodic features were extracted from the three vowels: F1, averaged over the inner 80% of the vowel, F2, averaged over the inner 80% of the vowel, pitch, averaged over the inner 80% of the vowel, vowel duration, shimmer, jitter, harmonics-to-noise ratio, overall intensity and spectral emphasis.

All measurements were performed with Praat [11]. All signals had a sample frequency of 16 kHz. Shimmer was extracted with the dda method and jitter with the ddp method, which are Praat's default methods. Pitch bottom and pitch ceiling adjustments for the various algorithms were set to 75 Hz and 600 Hz, respectively. In our experience, especially the pitch bottom adjustments have a strong influence on the measured values as the inclusion of creaky voice stretches depends heavily on the pitch bottom value.

Spectral emphasis was calculated with a method described in [12]. The signal was low-passed filtered with a cut-off frequency of 1.5 times the mean fundamental frequency. The intensity of this low pass-filtered signal was then subtracted from the intensity of the unfiltered signal. The result of this calculation reflects the amount of energy in the higher harmonics of the signal.

3. Results

All eight parameters were analysed with mixed ANOVAs with register (CDS vs. ADS) and phrase position/vowel quality (touch, snake, fish) as within subjects-factors and maternal role (mother vs. non-mother) and task sequence (first ADS, then CDS or vice versa) as between subjects-factors. For the purpose of the present study, the effects of register and maternal role are the more interesting ones. Therefore, only the effects of these factors will be reported here. Note that the degrees of freedom are slightly different for the different parameters due to missing values.

F0 showed a main effect of register ($F(1,43) = 6.2$, $p < .05$), indicating higher pitch on all vowels in CDS. The same was true for F1 ($F(1,46) = 21.4$, $p < .001$), F2 ($F(1,46) = 8.6$, $p < .01$), and vowel duration ($F(1,46) = 45.6$, $p < .001$). Neither f0 nor the formants nor vowel duration showed an effect of role, or an interaction between role and register (see Figures 1, 2, 3, and 4).

The picture was different for the source parameters. While jitter showed no significant within or between-subjects effects, shimmer showed a main effect of role ($F(1,37) = 7.8$, $p < .01$) (see figure 5). Both the effect of register as well as the interaction failed to reach significance (p's = .1) indicating a trend for the mothers to show higher shimmer in CDS, compared to the non-mothers. Similarly, for HNR, we found a significant effect of role ($F(1,45) = 4.3$, $p<.05$) indicating higher HNR in the non-mothers compared to the mothers, but no effect of register and no interaction (see figure 6). There were no effects for spectral emphasis.

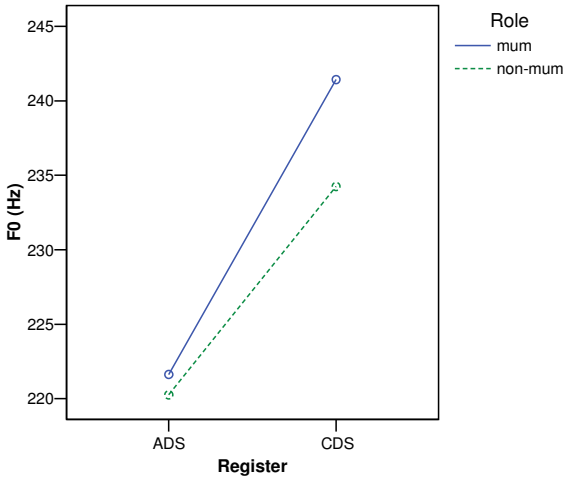

Figure 1: Fundamental frequency (f0) as a function of speech register and maternal role.

4. Discussion

Our findings show that both mothers addressing their children and non-mothers addressing an imaginary child increased their pitch, reduced their speech rate and in increased the first and second formant, adjustments that typically have been reported for CDS. This suggests that a range of parameters which are in service of linguistic and

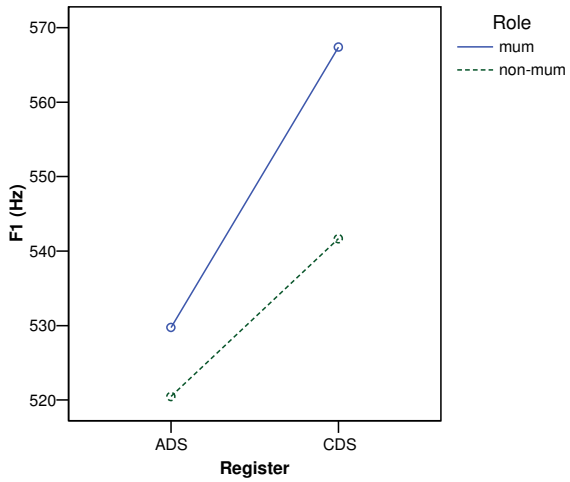

Figure 2: First formant (F1) as a function of speech register and maternal role.

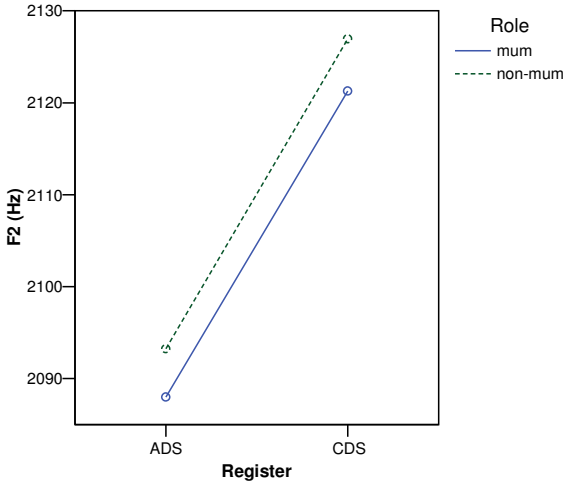

Figure 3: Second formant (F2) as a function of speech register and maternal role.

communicative function (such as speech prosody or hyperarticulation) are employed for both posed and spontaneous CDS.

Crucially, however, two of the measured source parameters, shimmer (perturbations in fundamental period amplitude) and harmonics-to-noise ratio (the relationship between periodic and non-periodic components in the speech signal) showed differences between mothers and non-mothers. This indicates that mothers exhibited more perturbations in amplitude, a finding that is in line with results from Trainor et al. [13] who showed that mothers exhibited more voice perturbations when singing to their babies than when singing in the absence of the child. Similarly, the lower harmonics-to-noise ratio in the mothers suggests that these speakers' voices exhibited more noise compared to the amount of harmonicity. This is also in line with results that report high degrees of breathiness in genuine CDS [14].

The design of our study does not yet allow us to exclude the possibility that the observed effects could also be linked to

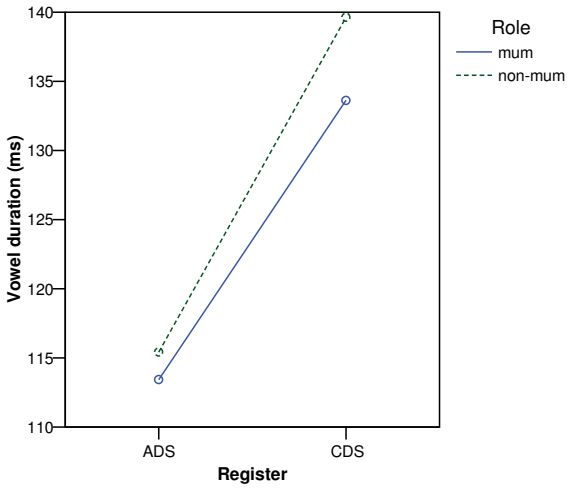

Figure 4: Vowel duration as a function of speech register and maternal role.

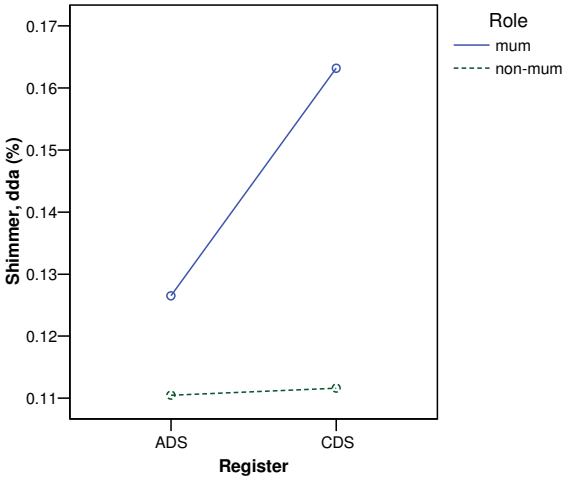

Figure 5: Shimmer (dda) as a function of speech register and maternal role.

the general presence or non-presence of an interlocutor. Therefore, additional experiments are currently under way where we will investigate the same vocal parameters for mothers and non-mothers interacting with non-kin children.

In sum, these results constitute preliminary evidence that voice quality parameters linked to periodicity and perturbations in the voice may serve as veridical indicators of genuine emotions experienced by speakers. Further research will have to confirm these findings, and to determine how the cortical and sub-cortical pathways for emotional and linguistic expressions differ. Furthermore, future research will address the specific mechanisms by which voice perturbations reflect arousal states associated with genuine emotions.

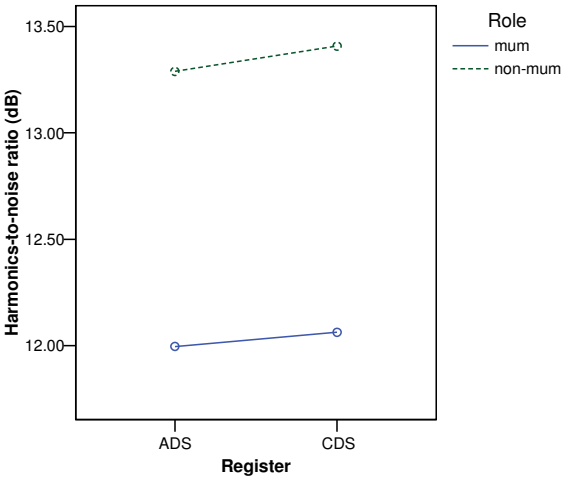

Figure 6: Harmonics-to-noise ratio (HNR) as a function of speech register and maternal role.

5. References

[1] Owren, M. J.; Bachorowski, J. A., 2003. Reconsidering the evolution of nonlinguistic communication: the case of laughter. *Journal of Nonverbal Behavior* 27(3), 183-2000.

[2] Argyle, M., 1996. *Bodily Communication.* London: Routledge.

[3] Borod, J. C., 1993. Cerebral mechanisms underlying facial, prosodic, and lexical emotional expression: A review of neuropsychological studies and methodological issues. *Neuropsychology 7 (4)*, 445-463

[4] Ziegler, W., 2003. Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech. *Aphasiology 17(1)*, 3-36.

[5] Hauser, M.D.; Akre, K., 1999. Asymmetries in the timing of facial and vocal expressions by rhesus monkeys: implications for hemispheric specialization. *Animal Behaviour*, 61, 391-400.

[6] Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227-256.

[7] Malatesta, C.Z.; Haviland, J.M., 1982. Learning display rules: The socialization of emotion expression in infancy. *Child Development*, 53, 991-1003.

[8] Nitschke, J.; Nelson, E.; Rusch, B.; Fox, A.; Oakes, T.; Davidson, R., 2004. Orbitofrontal cortex tracks positive mood in mothers viewing pictures of their newborn infants. *Neuroimage*, 21, 583–592.

[9] Jacobson, J. L.; Boersma, D. C.; Fields, R. B.; Olson, K. L., 1983. Paralinguistic features of adult speech to infants and small children. *Child Development,* 54 (2), 436-442

[10] Biersack, S., Kempe, V., Knapton, L. (2005). Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech. In: *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon.

[11] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[12] Heldner, M.; Strangert, E.; Deschamps, T. (1999). A focus detector using overall intensity and high frequency emphasis. In *Proceedings of the International Congress of Phonetic Sciences 99*, San Francisco, 1491–1493.

[13] Trainor, L.J.; Clark, E.D.; Huntley, A. & Adams, B.A., 1997. The Acoustic Basis for Infant-Directed Singing. *Infant Behavior and Development 20 (3)*, 383-396.

[14] N Campbell, N.; Mokhtari, P., 2003. Voice Quality: the 4th Prosodic Dimension, In *Proceedings of the International Congress of Phonetic Sciences 2003*, Barcelona.