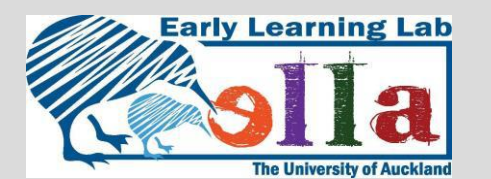


Caregiver and Infant Speech Matters:

An acoustic analysis on joint engagement during caregiver-infant peekaboo play in different digital contexts

Wan-Ting Yeh,
Alecia Moser,
Annette M. E. Henderson

School of Psychology, University of Auckland



Introduction

- Caregiver-infant interactions are fundamental to development. Early interactions involve social routines such as peekaboo (Tuomela, 2000).
- Peekaboo is universal gameplay that infants as young as four months engage in with the help of repetitive behaviours and caregivers' speech (Rochat et al., 1999).
- Caregiver vocalisation signals joint engagement during caregiver-infant interactions (Strouse et al., 2013). For example, caregivers use speech cues to help infants play peekaboo (Montague & Walker-Andrews, 2001) and use rising contours when

- interacting with less responsive infants (Niwano & Sugai, 2003).
- Infants vocalise to communicate their affective state and seek attention from their caregivers (Kitamura & Burnham, 1988).

Gap:

In today's world, caregiver-infant interactions are not restricted to face-to-face contexts but increasingly involve digital devices. The extent to which the acoustic features of verbal communication influence joint engagement during dyadic interactions with different digital devices remains unknown.

How does the association between **dyadic acoustic cues in verbal speech and joint engagement** during caregiver-infant **peekaboo** play differ across varied **digital** interactions?

Hypotheses:

- Caregivers' directional intonations provide informative cues (e.g., attention-elicitation) corresponding to the JCE score.
- Due to biological constraints in infant development, infants are most likely to respond to the social context with their voice volume and utterance duration.
- More salient acoustic features are expected to be found in FF and these features are more applicable for predicting JCE scores.

Significance:

Understanding the potential impact of **digital devices** on the **acoustic** properties is key to identifying how to preserve early dyadic **cooperative** interactions in our digital world.

Method

Participants

- 67 caregiver-infant dyads (37 male infants; 6 male caregiver)
- Infants aged between 18 to 26 months ($M_{age} = 21.4$ months, $SD = 2.20$).

Design & Procedure

- Dyads played peekaboo in two conditions: (1) face-to-face (FF): with a shared blank tablet present (2) video chat (VC): dyads interacted through separate devices
- Peekaboo play lasted 45 s per condition.

Joint Coordinated Engagement (JCE)

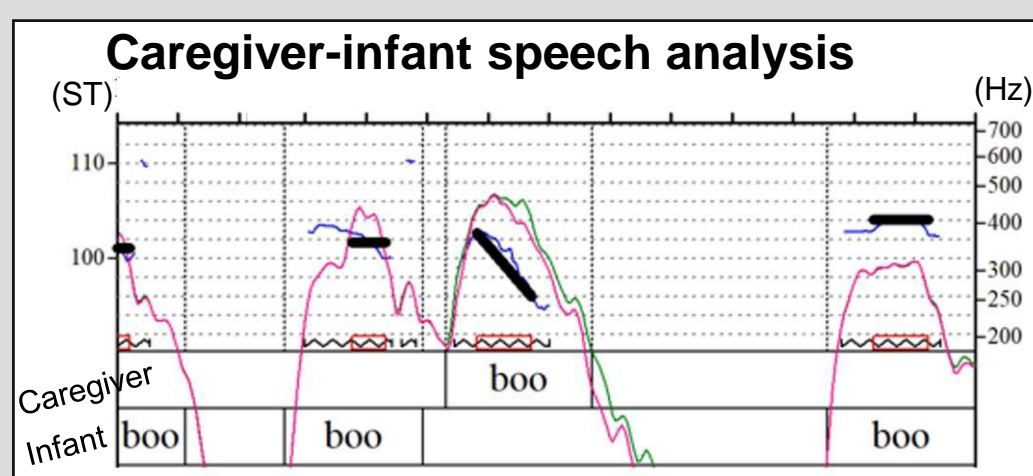
- Holistic measure of cooperation: joint attention, actions & communication (1 = None, 6 = High).
- Coded offline by two trained coders.

Speech Analysis

- 4,686 utterances were analysed across conditions ($M_{caregiver} = 1828$, $SD_{caregiver} = 58.7$; $M_{infant} = 515.5$, $SD_{infant} = 122.4$).
- Pitch: fundamental frequency (F0, Hz).
- Intonation: proportion (%) of syllables with large pitch movement (distance ≥ 4 ST).
- Speech time: duration of an utterance (s).
- Intensity: average SPL in dB.
- Analysis: *Praat* (Boersma & Weenink, 2021)

Statistical Analysis

- Linear regression (Predictors: dyadic acoustic variables & conditions; Outcome: JCE score)
- Analysis: *R* (R Core Team, 2014)

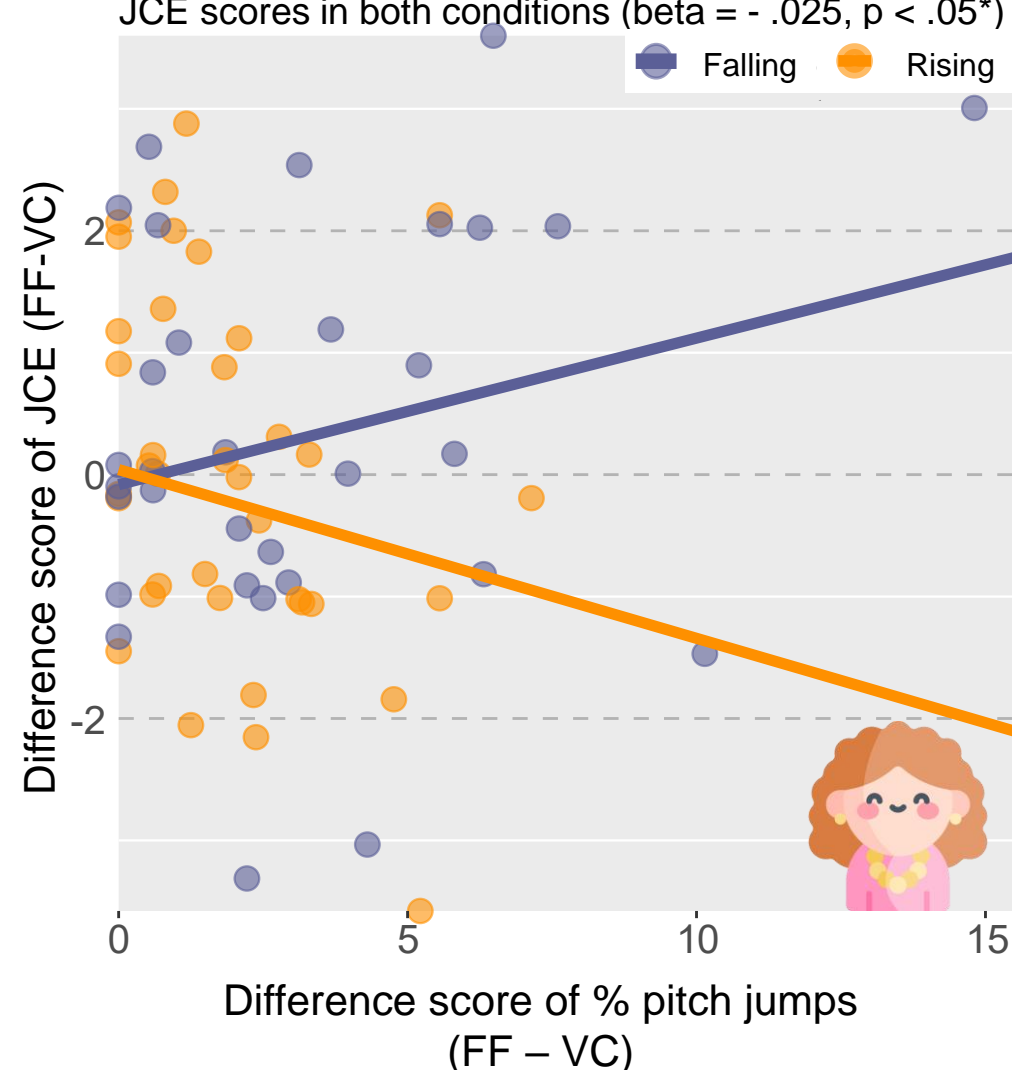


Results

Caregivers:

Directional intonation predicted JCE

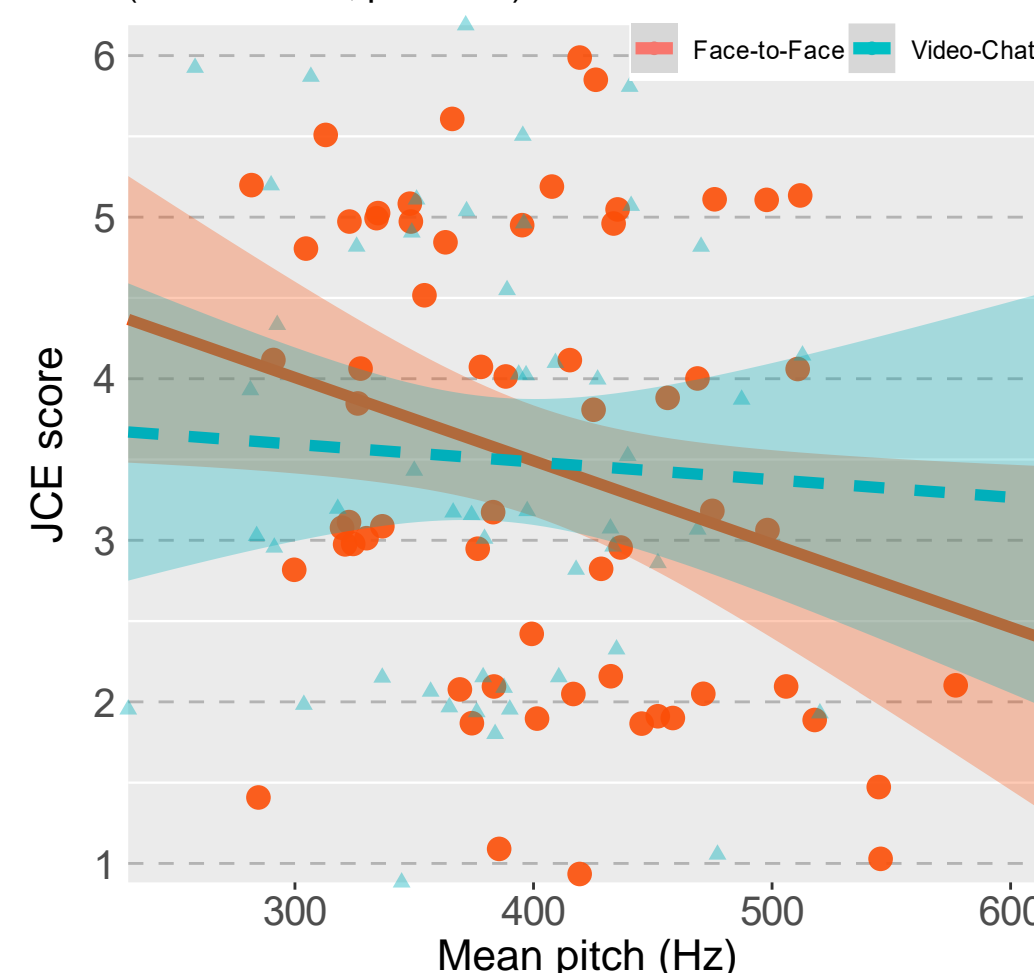
Dyads in which caregivers' directional intonation influenced JCE scores in both conditions ($\beta = -.025$, $p < .05^*$)



Infants:

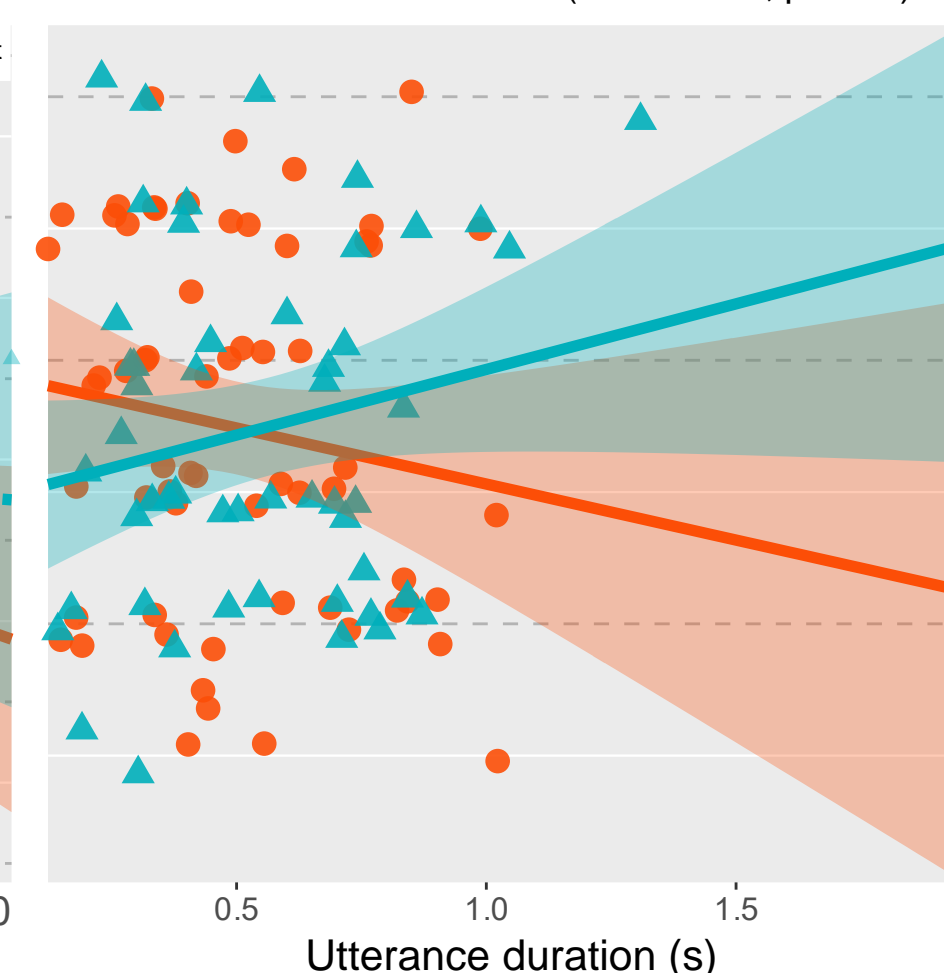
Higher pitch received lower JCE in FF

Dyads in which infants' mean pitch in FF: ($\beta = -.005$, $p < .01^{**}$)



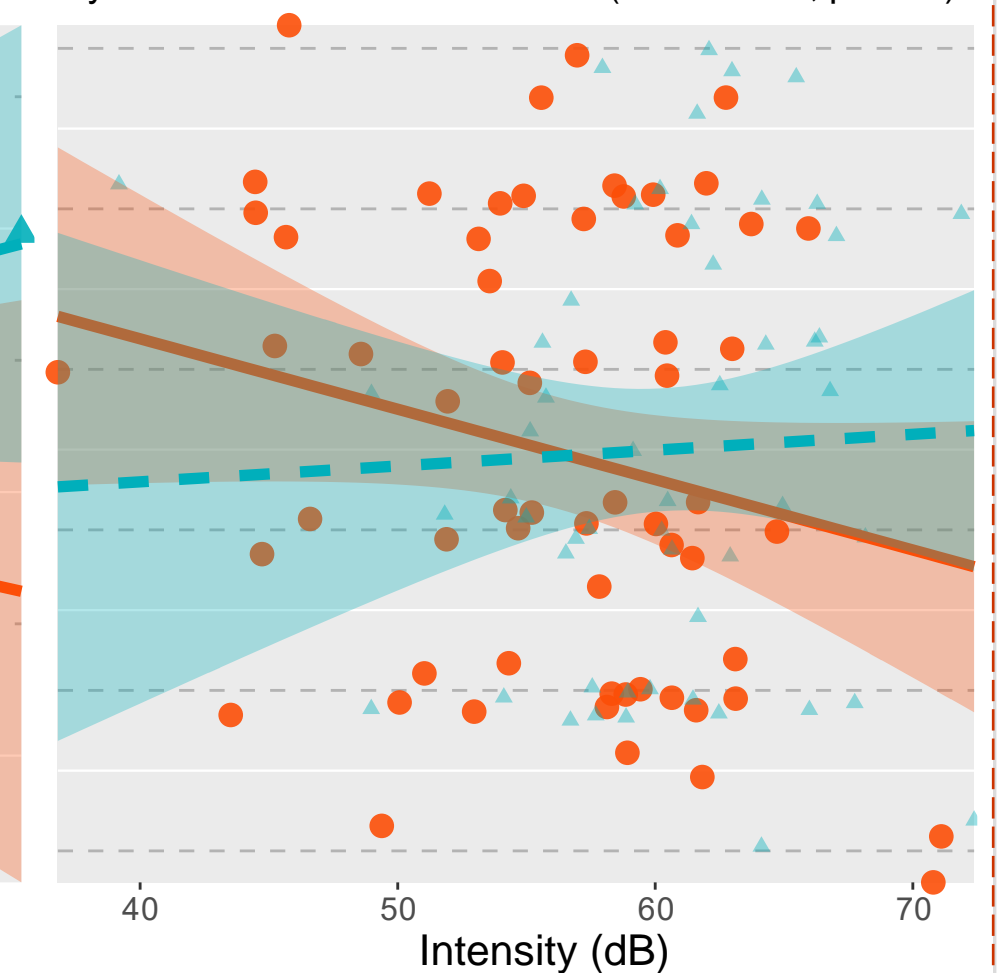
Utterance duration marginally predicted JCE

Utterance duration x Condition ($\beta = 1.81$, $p < .07$)



Loudness moderately predicted JCE in FF

Dyads in which infants' loudness ($\beta = -.05$, $p = .06$)



Figures: plotted with *ggplot2* package (Wickham, 2016)

Discussion

Parents' directional tones and infants' speech time predicted dyadic JCE during peekaboo in both digital interactions.

Infants' higher mean pitch and volume corresponded to lower JCE in the shared tablet face-to-face, but not the video-chat interaction.



Wan-Ting Yeh
PhD Candidate
wyeh841@aucklanduni.ac.nz

Caregivers' directional intonations

- Rising tone** → lower JCE: Rising tones are associated with attention elicitation (Niwano & Sugai, 2003).
- Falling tone** → higher JCE: Falling tones indicate statements and instructions (Show & Balog, 2002).

Infants' pitch, duration & intensity

- Higher pitch and loud voice** → lower JCE in FF: Pitch and volume are correlated with attention-seeking in face-to-face interactions (Hsu et al., 2014).
- Longer utterance** → higher JCE in VC: This might emphasise the importance of verbal cues in a physically separated setting.

Acknowledgments

Special thanks to the families who participated in this study, the ELLA team, and our founders.

Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.53, retrieved 8 September 2021 from <http://www.praat.org/>
Hsu, H. C., Iyer, S. N., & Fogel, A. (2014). Effects of social games on infant vocalizations. *Journal of Child Language*, 41, 132-154.
Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1), 85-110.
Mertens, P. (2014). Polytonia: A system for the automatic transcription of tonal aspects in speech corpora. *Journal of Speech Sciences* 4 (2), 17-57.
Montague, D. P., & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology*, 37(6), 826-838.
Niwano, K., & Sugai, K. (2003). Maternal accommodation in infant-directed speech during mother's and twin-infant vocal interactions. *Psychological Reports*, 92(2), 481-7.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology*, 35(4), 950-957.
R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025-1058
Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective co-viewing: Preschoolers' learning from video after a dialogic questioning intervention. *Developmental Psychology*, 49(12), 2368-2382.
Tuomela, R. (2000). *Cooperation: A philosophical study*. Dordrecht: Kluwer.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer, NY.

