

# The correlation underlines multimodal communication: investigating the trade-off and communicative efficiency hypotheses

Yanran Zhang<sup>1</sup>, Yan Gu<sup>2</sup>

<sup>1</sup>Tilburg University

<sup>2</sup>University of Essex & University College London

y.zhang@tilburguniversity.edu, yan.gu@essex.ac.uk

Language production is complicated because it is accompanied by prosodic features, linguistic speech, and gestures. Speakers can modulate their speech (e.g., pitch, speaking rate or words) and gestures according to adult (adult-directed language) (Aylett & Turk, 2006; Pluymaekers et al., 2005) or child (child-directed language) addressees (Campisi & Özyürek, 2013; Soderstrom et al., 2008). However, we know little about the exact relationship between different aspects of language production and how they are combined together to contribute to successful language communication according to different listeners. The tradeoff hypothesis assumes that the more complex the speech is or the larger effect in speech to convey the meaning, the higher the possibility of a gesture occurring to take over some of the communicative load (de Ruiter, Bangerter, & Dings, 2010). However, the communicative efficiency hypothesis proposes that speakers should make more effort to prevent communication failures when communication understandability is at risk, which means they will not minimize their production effort but make more effort to maximize understandability through conversation (e.g., increasing prosodic saliency and gesture rate simultaneously). In a live programme, broadcasters are trained to achieve communication understandability in a limited time. If the content requires less effort to achieve communicative success, broadcasters can speak faster, and produce fewer gestures. Alternatively, they may speak slower and produce more gestures to increase the likelihood of communicative success even if it requires greater temporal and manual effort. In addition, it could also be the case that broadcasters allocate more weight to certain modalities to achieve communicative efficiency as predicted by the tradeoff hypothesis.

To test whether there is a consistent adjustment in different modalities when addressing different audiences and whether the production of these multimodal adjustments is carried out independently in each modality (prosody, linguistic and gestures), we investigated how broadcasters organized their multimodal language production on both a regular adult-directed broadcasting (ADB) and a child-directed broadcasting (CDB) programme. Moreover, we studied if different audiences (children vs. adults) modulated the adoption of communicative efficiency and effort. Forty-six future broadcasters produced live programmes in which they explained four pictures to an adult and child audience respectively (within-subject). First, we compared multimodal production as a function of programmes to show how children audiences will affect communication efficiency. We computed a hyper score by computing a ratio between the values of ADB and CDB ( $ADB/CDB, >1$  means hyper adjustment) for each cue to show the degree of adjustment. Second, within a programme, we analysed the correlations between prosodic (F0, intensity, speaking rate), linguistic (lexical diversity, word frequency) and gestural (gesture types, gesture rate, duration) cues to see whether multimodal cues go hand in hand or are complementary to each other.

The findings reveal that: First, broadcasters had a general adaptation for children audiences for most cues except speaking rate ( $p=.65$ ). In addition, the degree of adjustment is not necessarily the same, but the mean intensity hyper-score correlated with the lexical diversity hyper-score, and the word frequency hyper-score predicted the mean duration of representational gestures hyper-score as shown in Figure 1. Furthermore, regardless of programmes, a faster speaking rate associated with simpler words (ADB ( $r(44) = .43, p=.003$ ), CDB ( $r(44) = -.31,$

$p=.04$ )), whereas the lexical diversity was negatively related to gesture rate (ADB ( $r(44) = -.38, p=.01$ ), CDB ( $r(44) = -.44, p=.002$ )). Mean intensity goes hand in hand with the average number of beat times (the repeated times of every beat during one beat gesture) (ADB ( $r(44) = .60, p<.001$ ), CDB ( $r(44) = .38, p=.009$ )), but representational gesture rate is negatively correlated with mean pitch ( $r(44) = -.47, p<.001$ ) and pause rate ( $r(44) = -.29, p=.03$ ) only in CDB. Moreover, there is no significant relationship between speaking rate and gesture rate in ADB ( $r(44) = -.03, p=.86$ ) or in CDB ( $r(44) = .12, p=.41$ ).

Although there is a general tendency of multimodal adaptation in CDB, there is also a trade-off between cues such as a higher lexical diversity relates to a lower gesture rate. However, the gesture rate was not affected by the speaking rate, and broadcasters applied communicative efficiency between intensity and average beat times, as both auditory stress and visual emphasis need to combine together to help their communication because of the time constrain. These results suggest that communicative efficiency and trade-off hypothesis may not underscore multimodal adjustments in CDL and ADL, as certain signal channels are not audience-oriented, whereas other channels may compensate for it.

**Index Terms:** multimodal communication, trade-off hypothesis, communicative efficiency hypothesis, broadcasting

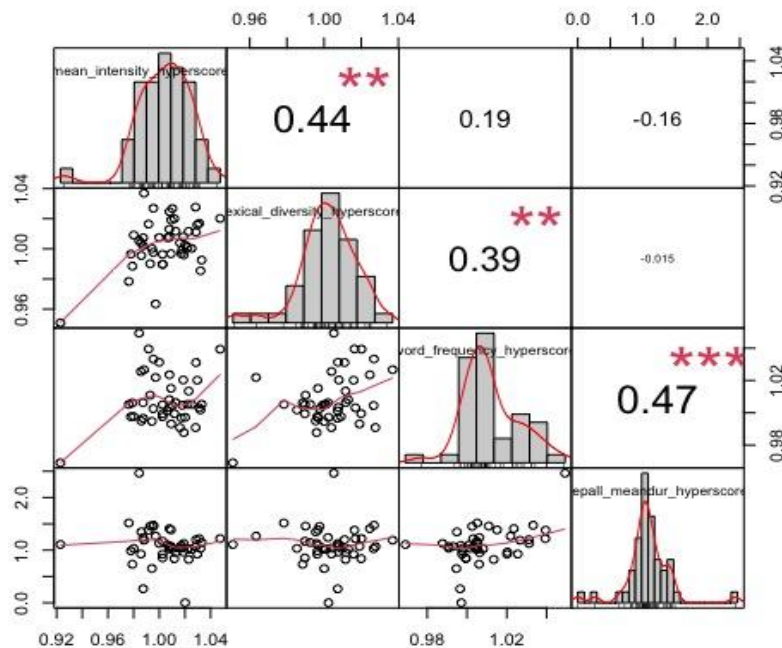


Figure 1: Correlation between mean intensity and lexical diversity, word frequency and representational gesture duration.

## References

- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119(5 Pt 1), 3048-3058.
- Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, 47(1), 14-27.

- de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4(2), 23-248.
- Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118(4), 2561-2569.
- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4), 869-902.