**Vocal gestures in early multimodal communication: A commentary on Karadöller, Sümer and Özyürek**

Karadöller and colleagues (2025) propose that "speaking children employ all mediums available to them to communicate their message" (p.23). The authors give evidence of infants' use of gestures to communicate in the absence of the appropriate lexical or syntactic repertoire, which may "lighten the cognitive load" (p.11) of early speech production. The authors' focus is on visual-spatial gestures produced mostly with the hands. Extending their account, here I propose to include what I refer to as *vocal gestures* in the speech-plus-gesture model.

Vocal gestures[[1]](#footnote-1) are relatively well-documented in case studies of language development, particularly in the early expression of onomatopoeia. Infants use prosodic, phonetic or phonological features to vocally express a meaning, rather than the target form. Elsen’s (1991) daughter Annalena, acquiring German, produces imitative forms created to reflect her own interpretation of environmental sounds. As reported in Laing (2014), her form for *Biene* "bee" constitutes a buzzing sound, *Vogel* "bird" a chirping sound with high pitch, and *Musik* "music" a sing-song voice. Elsen (1996) reports Annalena’s use of a hoarse voice to express *Krähe* "crow", while productions of *Hase* "hare" and *Schwein* "pig" involve "snuffling" and "grunting", respectively. Similarly, at 19-22 months, Hildegard Leopold, acquiring English and German, produces /kx/ to indicate something "tastes bad" (1939, p.140). Hilde Stern, acquiring German, produces "f-f-f" to indicate a matchstick, deriving from the action of blowing out a match, and later extended to candle, lamp and "other objects of light" (Stern & Stern, 1928, in Werner & Kaplan, 1963). Hilde's sibling, G, produced "*ö-ö-ö* pronounced rhythmically, with effort…signifying the strain of the horses involved in the pulling of the car" (*ibid*, p.102).

In all these examples, infants recruit additional resources, beyond those of target-like speech, to communicate meaning. Vocal gestures are not in the visual-spatial modality (though it is likely that facial expressions are involved, e.g. when "snuffling" to represent *hare*), yet they align with Karadöller et al.'s argument that gestures serve as an alternative means of communication when production ability is limited. This is a point made specifically regarding onomatopoeia in the infant production literature (Kauschke & Klann-Delius, 2007; Laing, 2019), but may be extended to incorporate iconicity more broadly.

What is striking about the examples above is that they occur as both vocalization and gesture. Annalena's productions of *Hase* "hare" and *Schwein* "pig" are not transcribed phonetically, but these forms are described in a way that implies facial gestures and vocal productions (‘snuffling, immediately recognizable as hare-like’ (p.209) and ‘grunting for a long time’ (p.312) Elsen, 1991; translation author’s own). Such forms recruit facial and perhaps manual (or other) movements to express a meaning, alongside a relevant vocalization that may include phonetic or prosodic features not typically found in the target language. Indeed, the distinction between vocalization and gesture may not be meaningful in early production: the gesture is part of the vocalization, and the vocalization may be meaningless without its accompanying gesture. In Karadöller and colleagues' account, iconic gesture production is reported to develop alongside its corresponding verb at 17-36 months (Özçalişkan et al., 2014). However, the examples of vocal gestures reported here are among the very first words that infants produce: in Elsen's (1991) account, 10 vocal gestures were produced before Annalena’s first birthday, including those reported above (see Laing, 2014 Table 1).

Werner and Kaplan (1963) examine the distinction between vocalization and other bodily movements, and describe the unfolding of language from the most basic reflexes. Their premise is that language is derived from non-linguistic forms, including bodily movements and vocalizations that they describe as "overdetermined"; that is, infants may adapt existing motor (including vocal) repertoires to new functions in order to express need. When it comes to early representations, there is no differentiation between vocalizations, gestures, and any other bodily movements; these are considered part of a "total organismic matrix" (p.77), which is gradually differentiated. This aligns with recent research by Borjon and colleagues (2024), showing that infants aged 9-24 months co-activate hand and head movements when vocalizing. This co-activation may be a crucial component of early speech-plus-gesture productions, including vocal gestures.

Imitative signals to indicate need may be the first vocal forms, before being used to represent meaning: eating, or the need for food, is a prime example of this signal-based vocal development, discussed in depth by Werner and Kaplan (1963). The authors suggest that sounds generated during ingestion become established by infants as a signal for requesting food (e.g., lip-smacking to indicate hunger, Bahorski et al., 2021), eventually expanding into a lexical unit to express food-related requests (e.g., /mam:a/ to request food, produced among Hildegard Leopold’s first five words; Leopold, 1939), and perhaps leading to a more conventional vocal form such as *mummy*. This example, which shows primordial vocalisations extending first to vocal gestures and then to formalised meaningful units in an infant’s vocabulary, is proposed as one of many referential expansions that take place during vocal development.

Acredolo and Goodwyn (1988) found that infants’ early use of symbolic gestures in reference to objects (e.g., "panting" to designate *dog* - which we might consider to be a vocal gesture - or arm-spreading to represent *aeroplane*) correlated with their lexical development, leading them to establish a 10-word vocabulary earlier than infants who did not use many such gestures. Crucially, 68% of gestures produced by infants were established independently (i.e. not learned from the caregiver). Of the 32% of gestures acquired from routines shared with the caregiver, around half of these were then spontaneously abstracted from the routine and applied in other contexts by the infant. Again, this suggests that expressing oneself through whatever means are available may be an important factor across development. Early gesture production predicts later vocabulary size, perhaps because gestures are an early expression of a child's ability to communicate: gesture use and vocabulary may be part of the same continuous process of production.

The limitation of this account is the lack of reported evidence, and formalized definition, of infants' vocal gestures. These are typically drawn from case studies, and rely on annotators to record unconventional productions that show consistent meaning. Vocal gestures may also include facial expressions or prosodic properties, which are difficult to record systematically and may be more variable in the infant's production. A move towards a speech-plus-gesture account of early word production should increase the data on infants' use of all kinds of gestures, including vocal gestures. This would be highly valuable for our understanding of early production: it is possible that in experimenting with unconventional productions, infants are mastering the structures and segments needed for target-like speech production.

Finally, iconicity plays a central role in vocal gesture production. All the examples referenced above draw on iconic expressions of a meaning; either vocalization or gesture or a combination of these lead vocal gestures to be highly iconic. This aligns with Karadöller and colleagues' account of spoken language acquisition, and while their account focuses on later acquisition, their claims regarding iconicity as a means of supporting communication when speech is limited remains consistent, right from the very first words. The authors note that "what children know and can express as they acquire language might be very limiting when one focuses only on speech" (p.23); to shift the focus to include the expressive vocal gestures that infants produce may tell us a lot more about how infants learn to use language.

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development 59*(2), 450-466.

Bahorski, J. S., Mumbower, R., & Pocchio, K. E. (2021). Describing Maternal Knowledge of Infant Feeding Practices. *Journal of Pediatric Health Care*, *35*(3), 278–284.

Borjon, J. I., Abney, D. H., Yu, C., & Smith, L. B. (2024). Infant vocal productions coincide with body movements. *Developmental Science*, *27*(4), e13491.

Elsen, H. (1991). *Erstspracherwerb: Der Erwerb des deutschen Lautsystems*. Deutscher Universitaets Verlag.

Elsen, H. (1996). Two routes to language: Stylistic variation in one child. *First Language*, *16*, 141-158.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (2025). First-language acquisition in a multimodal language framework: Insights from speech, gesture, and sign. *First Language*.

Kauschke, C., & Klann-Delius, G. (2007). Characteristics of maternal input in relation to vocabulary development in children learning German. In I. Gülzow & N. Gagarina (Eds.), *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept* (pp.181–204). De Gruyter Mouton.

Laing, C. (2014). A phonological analysis of onomatopoeia in early word production. *First Language*, *34*(5), 387–405.

Laing, C. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition*, *11*(2), 173–187.

Leopold, W. F. (1939). *Speech development of a bilingual child, 1: Vocabulary growth in the first two years*. Northwestern University Press.

Özçalişkan, Ş., Gentner, D., & Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children’s early verbs? *Applied Psycholinguistics*, *35*(6), 1143–1162.

Werner, H., & Kaplan, B. (1963). *Symbol Formation*. John Wiley & Sons.

1. Werner and Kaplan (1963, p.101) refer to such forms as "vocal depictive forms", and Stern & Stern (1928, cited in *ibid*) as "designatory vocal imitations". [↑](#footnote-ref-1)