

NOTICE: this is the author's version of a work that was accepted for publication in *Current Biology*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication.

A definitive version was subsequently published in *Current Biology*, 2012 Mar 6; 22(5):R160-1. doi:10.1016/j.cub.2012.01.032

Dispatch

Speech Development: Toddlers Don't Mind Getting it Wrong

Piers Messum

112 Warner Road, London SE5 9HQ, UK.

p.messum@ucl.ac.uk

Ian S. Howard

Computational & Biological Learning Laboratory, Department of Engineering,
University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK.

ish22@cam.ac.uk

A recent study has found that toddlers do not compensate for an artificial alteration in a vowel they hear themselves producing. This raises questions about how young children learn speech sounds.

When adults hear themselves over headphones, they compensate for experimental manipulation of the qualities of their vowels [1]. Given that children are widely believed to learn speech sounds by imitation [2,3], one would expect toddlers, too, to be monitoring their output. However, MacDonald and colleagues [4] have recently reported in *Current Biology* that when toddlers intend to say "bed", they seem indifferent to their output sounding like "bad".

MacDonald *et al.* [4] tested the self regulation of speech by adults, young children (mean age 4 years 3 months) and toddlers (mean age 2 years 6 months). The younger subjects played a video game, where a character moved in response to the 'magic' word "bed"; adults simply said the word. All spoke into a microphone, wore headphones, and heard their speech amplified and mixed with noise to mask bone-conducted feedback. After 20 utterances of "bed", the speech signal was manipulated by moving the first and second formants up by 200 Hz and down by 250 Hz respectively. If the subjects maintained their original articulation, what they heard would now sound like "bad" for the next 30 utterances.

The adults partially compensated for the manipulation, changing their articulation and stabilising their output after about 10 utterances. The young children also compensated appropriately, albeit less reliably. The toddlers were indifferent to the change in what they were hearing. None made any compensatory shifts.

This last result appears to be inconsistent with the widespread notion that children learn speech sounds by imitation — that is, by self-supervised auditory matching using self-developed criteria of sound similarity. Such a strategy would require of them that they both attend to their own output over the protracted period during which L1 pronunciation develops and act on any discrepancy they notice between their output and targets internalised from the ambient language. In this experiment, the toddlers didn't behave as this would predict.

MacDonald *et al.* [4] discuss reasons why their data might not be representative of toddlers' normal behaviour. These will all need further examination. However, the authors identify two main questions that arise assuming their results do generalise: first, why is the development of self regulation of speech production using auditory feedback delayed, rather than being present from the start of speech? And second, how does early vocal learning take place in its absence?

To answer the second question, MacDonald *et al.* [4] point to the series of experiments conducted at Indiana and Cornell, where the discovery of non-vocal tutoring of young male cowbirds learning to sing by non-singing females has been extended into evidence of a similar paradigm operating in early child development. In a number of experiments with infants, the responsiveness of social partners to immature behavior has been shown to be perceived and used by the young learners to generate more advanced forms of vocalization [5,6].

For this paradigm to account for speech sound development, social partners must play their part. Importantly, Pawlby [7] found that in imitative vocal exchanges between mothers and their infants, it was the mothers who imitated their children more than 90% of the time. These results have been confirmed in subsequent studies. Mothers reflect (or mirror) what their children say, but such imitation generally takes the form of reformulation into well-formed sounds of the ambient language, rather than simple mimicry (as would occur in impersonation). So infants are presented with the linguistic interpretation of what they have done immediately after they do it, one favourable condition for associative learning [8].

The earliest proposal of socially guided vocal learning leading to speech that we are aware of was made by the educationalist Caleb Gattegno in 1962 [9]. He noted that as soon as a baby produces sounds, someone in the environment starts imitating the baby. He asserted that, "It is the imitation by other people of some of the sounds produced by babies that channels the production of some sounds of the mother tongue. This is not learning of what exists [i.e. imitation], but agreeing to separate a set of noises among all possible noises because of the feedback that the language environment provides." At this stage, no imitation of speech sound qualities on the part of the infants is involved: "Production of sounds being spontaneous, and similarity of these sounds with those of the environment being only approximate, babies do not feel the compulsion to alter their own activity to agree with an outer criterion." He went on to describe how speech sound equivalences create the 'bridge' to learning the pronunciation of words.

Gattegno's insights have been examined in the wider context of child and adult speech data [10] and his model of speech acquisition developed further and tested with caregiver reformulations as input to an infant computer model that learns the pronunciation of words [11,12]. Independently, the Asada group in Osaka [13,14] have also used reformulation/mirroring by a caregiver to train vowel qualities in a physical vocal tract model.

Returning to the first question identified by MacDonald *et al.* [4], of why self monitoring is delayed, it is not surprising within this new paradigm that reliance on auditory information for self monitoring comes late in a child's speech development. The starting point for speech production is motor exploration and the proposal is that an infant has no early need to reconceive his speech sounds in auditory terms in order to compare and evaluate his production with that of others. As auditory feedback is then only a secondary sensory information source for speech sounds, its use will develop accordingly. For haptic and spatial information, Gori *et al.* [15] recently found that one sense dominates totally in tests of multisensory integration in children up to 8 years of age. Reviewing this and recent papers reporting similar results, Ernst [16] said that it is unclear why integration emerges so late, but argued that it is unlikely to only be the result of the challenges caused by growth and sensory reorganisation. Whatever the reasons, young children do ignore sensory data that they do not consider to be primary.

Children who are usually a little older than the toddlers tested by MacDonald *et al.* [4] have often been reported to persist with the pronunciation of an incorrect word form, even when they deploy the speech sound they need elsewhere. The phenomena are discussed under various labels: "fis/fish", "puzzle/puddle/puggle", "guck" for "duck" (persistently), and so on. The puzzle is that the child hears adult speech correctly, but not, it seems, his own. Out of a range of hypotheses addressing this (summarised in [17,18]), none conclusively explains the whole range of situations where children are apparently oblivious to the reality of what they are saying. MacDonald *et al.*'s [4] results suggest that these behaviours may not be the manifestation of a novel absence of attention by a child to his own output, but a continuation of what is systematic in the behaviour of toddlers.

References

1. Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science* 279, 1213-1216.
2. Fry, D.B. (1968). The phonemic system in children's speech. *British Journal of Disord. Commun.* 3, 13-19.
3. Kuhl, P.K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci. USA* 97, 11850-11857.

4. MacDonald, E.N., Johnson, E.K., Forsythe, J., Plante, P., and Munhall, K.G. (2012). Children's development of self-regulation in speech production. *Curr. Biol.* 24th January issue.
5. Gros-Louis, J., West, M.J., and King, A.P. (2010). Comparative perspectives on the missing link: communicative pragmatics. In *The Oxford Handbook of Developmental and Comparative Neuroscience*. M.S. Blumberg, J.H. Freeman, S.R. Robinson, eds. (OUP), pp. 684-707.
6. Goldstein, M.H., and Schwade, J.A. (2010). From birds to words: perception of structure in social interactions guides vocal development and language learning. In *The Oxford Handbook of Developmental and Comparative Neuroscience*. M.S. Blumberg, J.H. Freeman, S.R. Robinson, eds. (OUP), pp. 708-729.
7. Pawlby, S.J. (1977). Imitative interaction. In *Studies in Mother-Infant Interaction*. H.R. Schaffer, ed. (London: Academic Press), pp. 203-223.
8. Heyes, C. (2010). Where do mirror neurons come from? *Neurosci. Biobehav. Rev.* 34, 575-583.
9. Gattegno, C. (1962). *Teaching Foreign Languages in Schools: the Silent Way*, 1st edn. (Reading: Educational Explorers.)
10. Messum, P.R. (2007). The role of imitation in learning to pronounce. (London University: PhD thesis.)
11. Howard, I.S., and Messum, P.R. (2007). A computational model of infant speech development. In *Speech and Computer (SpeCom XII)*. (Moscow: Moscow State Linguistics University), pp. 756-765.
12. Howard, I.S., and Messum, P.R. (2011). Modeling the development of pronunciation in infant speech acquisition. *Motor Control* 15, 85-117.
13. Yoshikawa, Y., Asada, M., Hosoda, K., and Koga, J. (2003). A constructivist approach to infants' vowel acquisition through mother-infant interaction. *Connection Science* 14, 245-258.
14. Miura, K., Yoshikawa, Y., and Asada, M. (2007). Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver's vowel categories. *Adv. Robotics* 21, 1583-1600.
15. Gori, M., Del Viva, M., Sandini, G., and Burr, D.C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694-698.
16. Ernst, M.O. (2008). Multisensory integration: a late bloomer. *Curr. Biol.* 18, R519-R521.
17. Locke, J.L. (1979). The child's processing of phonology. In *Child Language and Communication: Minnesota Symposium on Child Psychology Volume 12*. W.A. Collins, ed. (Hillsdale, NJ: LEA), pp. 83-119.
18. Alvater-Mackensen, N., and Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua* 120, 1898-1909.