Commentary

Psychological Science
2015, Vol. 26(12) 1975–1977

DOI: 10.1177/0956797614563766
pss.sagepub.com

Do We Know What We're Saying? The Roles of Attention and Sensory Information During Speech Production

Sophie Meekings, Dana Boebinger, Samuel Evans, César F. Lima, Sinead Chen, Markus Ostarek, and Sophie K. Scott

Institute of Cognitive Neuroscience, University College London

Received 5/19/14; Revision accepted 11/20/14

Lind, Hall, Breidegard, Balkenius, and Johansson (2014a, 2014b) recently published articles tackling a core question concerning speech production: At which stage of processing are communicative intentions specified? Taking a position contrary to dominant models of speech production (e.g., Levelt, 2001), they suggested that utterances "often are semantically underspecified" (Lind et al., 2014a, p. 8) before articulation, and that "auditory feedback" (Lind et al., 2014a, 2014b) is an important mechanism for specifying the meaning of what one says. In their experiment (Lind et al. 2014b), they used real-time speech exchange, in which auditory information was occasionally manipulated while participants performed a Stroop task (i.e., naming the colors in which color names were printed). This methodology created trials in which participants produced the correct response (i.e., named the color of the text) while simultaneously hearing over headphones a recording of their own voice producing the incorrect response (i.e., reading the text). On these trials, when asked what they had said, participants sometimes reported the word that they had heard, rather than the word that they had produced. Failure to detect these intrusions was interpreted as evidence that auditory feedback is a "prime channel" (Lind et al., 2014a, p. 2) for monitoring the meaning of spoken words. We suggest that (a) the authors' data constitute evidence against a prime role for auditory feedback in monitoring the meaning of self-produced speech, (b) the experimental manipulation is not necessarily a manipulation of auditory feedback, and (c) the findings may be explained by task demands rather than by mechanisms of speech production.

In the experiment (Lind et al., 2014b), participants always heard their own responses, via normal bone conduction and also via headphones. In a subset of trials (4 out of 250 per participant), the word they heard over headphones was a recording, from earlier in the experiment, of the word that they currently needed to inhibit. Lind et al. focused their analysis and interpretation exclusively on those trials within this subset in which participants did not notice that the word had been exchanged. However, according to the criteria Lind et al. used, there was evidence that the majority of these exchanges were in fact detected (~73%), and only a minority went unnoticed (~27%). The frequent detection of the manipulations suggests that auditory feedback is unlikely to be a prime mechanism (Lind et al., 2014a) by which spoken intentions are specified, as participants must have had some awareness of what they actually said when they detected the exchanges.

Although the data seemingly contradict the authors' claim that speakers "always use [auditory feedback] as a source of evidence in interpreting their own utterances" (Lind et al., 2014b, p. 1199), it remains possible that speakers use auditory feedback some of the time, when self-monitoring. However, we argue that the experience of hearing the audio recordings differed so much from the usual experience of hearing one's own voice that it cannot be assumed to necessarily reflect the same processes. People who listen to a recording of their own voice notice that their voice sounds different from how they hear it when they speak; they primarily hear their own voice via bone conduction but other people's voices through air conduction. Lind et al. accounted for some of the acoustic differences between self- and other-produced speech as it is heard at the ear (e.g., by low-pass filtering the speech), but made no attempt to address

Corresponding Author:
Sophie K. Scott, University College London–Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, United Kingdom
E-mail: sophie.scott@ucl.ac.uk

others (e.g., by mimicking the spatial location of participants' own voices).

Furthermore, Lind et al. could not eliminate the perception of the concurrent somatosensory and bone-conducted auditory information that always results from speaking aloud. Within the dorsolateral temporal lobes, the neural response to self-produced speech differs from the response to other-produced speech (Agnew, McGettigan, Banks, & Scott, 2012; Wise, Greene, Buchel, & Scott, 1999), and peripherally, the loudness of one's voice is dampened by the stapedius reflex, which occurs when one speaks. It is critical to determine, therefore, whether the auditory feedback in the experiment reflected the auditory, somatosensory, and neural consequences of perceiving one's own voice while speaking, or whether it simply functioned more generally as an auditory distractor. This distinction matters, because there are many instances in which hearing other-produced speech influences perception and production. Properties of the speech sounds in the environment can be assimilated into a speaker's output (Delvaux & Soquet, 2007; Pickering & Garrod, 2004), and speech that a person is trying to ignore can be confused with attended speech, particularly if the utterances are semantically related (Brungart, 2001; Gray & Wedderburn, 1960). Thus, had participants heard any voice (not necessarily their own), they might have made the same misattributions. To ensure that the observed effect was dependent on participants hearing their own voice, the authors would have needed an additional control condition in which participants heard a voice other than their own overlaid on their vocal responses.

Finally, we question the extent to which performance on the Stroop task generalizes to natural speech interactions. Stroop interference results from competition between the color of the text and the distractor (the incongruent written word; MacLeod, 1991; van Veen & Carter, 2005). Both are automatically processed and prepared for response production, and executive-control systems are required for the final response selection. Although manipulation of stimulus onset asynchrony shows that Stroop interference dissipates over relatively short latencies (~400–500 ms; for reviews, see MacLeod, 1991; Roelofs, 2003), congruency sequence effects suggest that cognitive control is sustained and adjusted over longer time scales (Egner, 2007; Kerns et al., 2004). Cognitive-control demands likely reduce capacity for error monitoring of one's own speech. Indeed, individuals make fewer self-repairs of their speech in dual-task paradigms than in single-task paradigms (Oomen & Postma, 2002) and show poorer pre-articulatory self-monitoring when naming under time pressure than when naming under no time pressure (Dhooge & Hartsuiker, 2012)—effects that are concordant with the notion that error monitoring is under central control and is capacity limited (Levelt, 1983). The data reported by Lind et al. (2014b) seem best explained by breakdowns in error monitoring consequent to the executive processing demands of the Stroop task, rather than by processes underlying speech production more generally.

To conclude, we are sympathetic to the wider concept that the act of speaking helps people construe their intended meanings in interactions. However, the articles by Lind et al. (2014a, 2014b) do not make a convincing case for placing auditory feedback at the heart of this mechanism.

Author Contributions

All the authors contributed to writing this manuscript.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

S. K. Scott, S. Evans, and D. Boebinger were supported by a Wellcome Trust Senior Research Fellowship (Grant No. WT090961MA), awarded to S. K. Scott.

References

Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2012). Articulatory movements modulate auditory responses to speech. *NeuroImage*, *73*, 191–199.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, *109*, 1101–1109.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*, 145–173.

Dhooge, E., & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language*, *66*, 163–176. doi:10.1016/j.jml.2011.08.004

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioural Neuroscience*, *7*, 380–390.

Gray, J. A., & Wedderburn, A. A. I. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, *12*, 180–184.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026. doi:10.1126/science.1089910

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences, USA*, *98*, 13464–13471.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014a). Auditory feedback of one's own voice is used for high-level semantic monitoring: The "self-comprehension" hypothesis. *Frontiers in Human Neuroscience*, *8*, Article 166. Retrieved from http://journal.frontiersin.org/article/10.3389/fnhum.2014.00166/full

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014b). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, *25*, 1198–1205.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*, 163–184. doi:10.1080/01690960143000010

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, *27*, 169–226.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *NeuroImage*, *27*, 497–504.

Wise, R. J. S., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*, 1057–1061.