

Authors of target article: Hermann Ackermann, Steffen R. Hage, Wolfram Ziegler

Word counts: Abstract (42), Main Text (998), References (349), Entire Text (1499)

Title: Voluntary and involuntary processes affect the production of verbal and non-verbal signals by the human voice.

Authors: Carolyn McGettigan¹ and Sophie Kerttu Scott²

Institutions: Royal Holloway, University of London and University College London

Mailing addresses: 1: Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK and 2: Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham TW20 0EX, UK

Institutional telephone numbers: 01784 443529 and 020 7679 1144

Email: Carolyn.McGettigan@rhul.ac.uk and sophie.scott@ucl.ac.uk

Homepages:

www.carolynmcgettigan.com and <https://sites.google.com/site/speechskscott/>

Abstract:

We argue that a comprehensive model of human vocal behaviour must address both voluntary and involuntary aspects of articulate speech and non-verbal vocalizations. Within this, plasticity of vocal output should be acknowledged and explained as part of the mature speech production system.

Main Text:

In their account of the neural systems supporting vocal expression in humans, Ackermann, Hage and Ziegler suggest that emotional and “attitudinal” aspects of prosody might influence the execution of speech via cross-talk between basal ganglia loops processing emotion, motivation, and speech motor programmes. It is problematic to claim one system for spoken language, plus one or more others for paralinguistic or non-linguistic signals that are then added together to make a finished product of fluent, emotionally inflected speech. The division between lateral motor cortex and other systems in the production of human vocal signals is not a simple one, and might be better characterized by the degree of *voluntary control* over the vocal tract rather than according to the type of signals generated. For example, patients who have sustained lateral cortical injuries disrupting the voluntary production of speech can still produce spontaneous and natural-sounding laughter and crying, and *swearing* (Van Lancker & Cummings, 1999). Thus, articulate speech – swear words – can be produced involuntarily. Similarly, non-verbal emotional vocalizations can be produced under voluntary control – social laughter is typically timed to occur at the end of linguistic phrases, during both

speaking and signing (Provine & Emmorey, 2006). Recent work using functional MRI to explore the neural underpinnings of laughter showed a considerable involvement of lateral sensorimotor systems in the production of laughter under varying amounts of voluntary control (Wattendorf et al., 2012).

In everyday spoken language, voluntary modulation of the way we speak plays an essential role in the intentional expression of mood, intentions and aspirations. Hawkins and Smith (2001) illustrate this with the English phrase “I do not know”, the pragmatic sense of which can vary dramatically depending on how the words are articulated (compare the casual manner of “I dunno” with the suggestion of irritation in “I... do... not... know!”). We recently investigated the neural correlates of voluntary modulations of spoken language by asking participants in an MRI scanner to perform spoken impressions of accents and impersonations of familiar individuals (McGettigan et al., 2013). The peak activations associated with deliberate changes to speaking style (compared with speaking in a “normal voice”) were found in the left anterior insula and inferior frontal gyrus. These areas are classically associated with the production of spoken language (Blank, Scott, Murphy, Warburton, & Wise, 2002; Dronkers, 1996), yet in this case the linguistic content of the utterances was kept constant across the different conditions of the experiment. It is difficult to assert that these voluntary aspects of speech production should, or *could*, be added to speech separately from the “digital” information bound up in the phonemes, syllables and words of a language. Our recent results suggest that this

kind of flexibility is an integral part of the planning and control of speech and voluntary vocal behaviour.

Not all vocal modulations can be added to speech in a controlled manner. Ackermann and colleagues argue that linguistic and emotional prosodic information, which they see as digital and analogue, respectively, are coordinated in the basal ganglia, as “otherwise these two inputs would distort and corrupt each other” (p. 6). It is reductive to draw boundaries between linguistic and paralinguistic aspects of vocal behaviour, particularly when considering the role of linguistic prosody in disambiguation (for example, the contrast between a question and a statement). Furthermore, it is certainly the case that emotional states do corrupt articulate speech, as is shown when a person tries to produce speech during a fit of laughter, when overcome with grief, or when feeling extremely nervous – here, the *voluntary control* of vocalization is compromised, and articulate speech is taken over by the physiological effects of emotion on the functions of the vocal tract (see Figure 1;(Levenson, 2003).

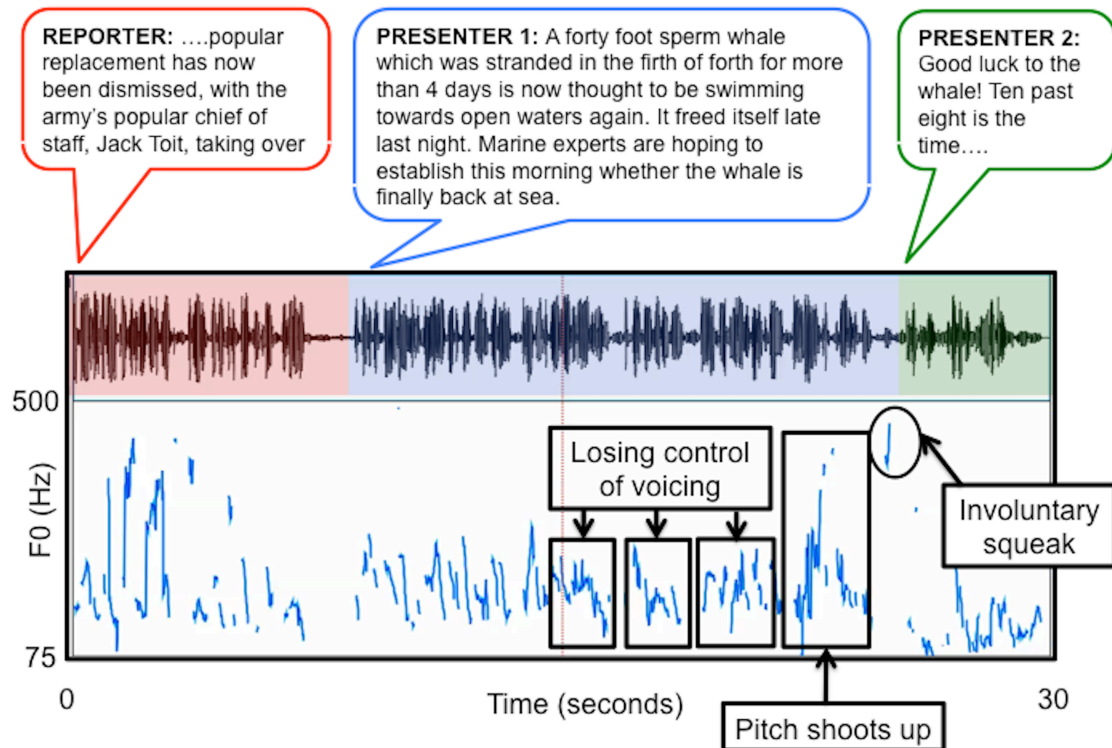


Figure 1. On live radio, Presenter 1 is amused by the Reporter's pronunciation of "Jack Toit". Although she manages to deliver her script, the pitch (F0) of her voice rises sharply as her emotional state constricts the vocal tract and renders her less able to control the source of the vocal signal (Ruch & Ekman, 2001).

Ackermann and colleagues claim that the basal ganglia might be essential for the acquisition of articulate speech during early childhood, while the behaviours of the mature speech production system are controlled by perisylvian cortical structures. There is evidence that the plasticity of vocal learning reduces in adolescence and adulthood, for example the marked persistence of first-language pronunciation in adult learners of a second language (J. E. Flege, MacKay, & Meador, 1999; J.E. Flege,

Yeni-Komshian, & Liu, 1999). However, speech can change in adulthood – one study showed that vowels in the speech of Queen Elizabeth II have, over several decades, gradually moved closer to the standard British English spoken by her subjects (Harrington, Palethorpe, & Watson, 2000). Similarly, there is extensive evidence for the recovery of speech in the adult system after stroke (Blank, Bird, Turkheimer, & Wise, 2003). It is difficult to estimate the extent to which these gradual changes to speech come about under conscious voluntary control. We continue to learn new information at all levels of the linguistic hierarchy throughout the lifespan, and the extent to which an individual changes their speech, voluntarily or not, can vary over both long and short timescales. With reference to the authors' proposal, we therefore pose the question: How do relearned and remapped behaviours in the adult speech production system fit within a model where the contributions of the basal ganglia end after childhood language acquisition?

We are encouraged by an approach to modelling human vocal behaviour that incorporates its social, emotional and linguistic aspects. However, we urge caution in attempts to divide the speech signal into distinct types of information served by specific underlying functional sub-systems. We argue that vocal behaviour is better characterized in terms of voluntary versus involuntary control of a complex motor act, regardless of its informational content. Further, given the evidence that vocal behaviour remains plastic and flexible into adulthood, we question the extent to which this plasticity need be mechanistically distinct from childhood language acquisition.

References

- Blank, S. C., Bird, H., Turkheimer, F., & Wise, R. J. (2003). Speech production after stroke: the role of the right pars opercularis. *Annals of Neurology*, *54*(3), 310-320. doi: 10.1002/ana.10656
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, *125*(Pt 8), 1829-1838.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159-161. doi: 10.1038/384159a0
- Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973-2987.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*, 78-104.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? Elizabeth II's traditional pronunciation has been influenced by modern trends. *Nature*, *408*(6815), 927-928. doi: 10.1038/35050160
- Hawkins, S., & Smith, R. (2001). Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics*, *13*, 99-188.
- Levenson, R. W. (2003). Blood, Sweat, and Fears: The Autonomic Architecture of Emotion. *Annals of the New York Academy of Sciences*, *1000*, 348-366.
- McGettigan, C., Eisner, F., Agnew, Z. K., Manly, T., Wisbey, D., & Scott, S. K. (2013). T'ain't what you say, it's the way that you say it – left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *Journal of Cognitive Neuroscience*. doi: doi:10.1162/jocn_a_00427
- Provine, R. R., & Emmorey, K. (2006). Laughter among deaf signers. *Journal of Deaf Studies and Deaf Education*, *11*(4), 403-409. doi: 10.1093/deafed/enl008
- Ruch, W., & Ekman, P. (2001). The expressive pattern of laughter. In A. W. Kaszniak (Ed.), *Emotion, qualia, and consciousness* (pp. 426-443). Tokyo: Word Scientific.
- Van Lancker, D., & Cummings, J. L. (1999). Expletives: neurolinguistic and neurobehavioral perspectives on swearing. *Brain Research Reviews*, *31*, 83-104.
- Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., & Celio, M. R. (2012). Exploration of the Neural Correlates of Ticklish Laughter by Functional Magnetic Resonance Imaging. *Cerebral Cortex*. doi: 10.1093/cercor/bhs094