Original Article

Impoverished encoding of speaker identity in spontaneous laughter

CrossMark

Nadine Lavan [a,b,*], Bethanie Short [a], Amy Wilding [a], Carolyn McGettigan [a,c]

[a] Department of Psychology, Royal Holloway, University of London, Egham, UK
[b] Department of Psychology, Brunel University, London, UK
[c] Institute of Cognitive Neuroscience, University College London, London, UK

A B S T R A C T

Our ability to perceive person identity from other human voices has been described as prodigious. However, emerging evidence points to limitations in this skill. In this study, we investigated the recent and striking finding that identity perception from spontaneous laughter - a frequently occurring and important social signal in human vocal communication - is significantly impaired relative to identity perception from volitional (acted) laughter. We report the findings of an experiment in which listeners made speaker discrimination judgements from pairs of volitional and spontaneous laughter samples. The experimental design employed a range of different conditions, designed to disentangle the effects of laughter production mode versus perceptual features on the extraction of speaker identity. We find that the major driving factor of reduced accuracy for spontaneous laughter is not its perceived emotional quality but rather its distinct production mode, which is phylogenetically homologous with other primates. These results suggest that identity-related information is less successfully encoded in spontaneously produced (laughter) vocalisations. We therefore propose that claims for a limitless human capacity to process identity-related information from voices may be linked to the evolution of volitional vocal control and the emergence of articulate speech.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Listeners are readily able to extract information about a speaker's identity from the human voice: Studies have shown that we can recognise (familiar) individuals from their voices (Mathias & von Kriegstein, 2014 for a recent review; Kreiman & Sidtis, 2011) and can successfully discriminate between (unknown) speakers (Reich & Duke, 1979; Van Lancker & Kreiman, 1987; Wester, 2012). How accurately and reliably we can extract these kinds of information depends on the task, listener characteristics and stimulus characteristics: for example, studies report that the duration of the test stimuli (Schweinberger, Herholz, & Sommer, 1997), the information encoded in the stimuli (Bricker & Pruzansky, 1966) as well as the retention interval between exposure and test (for recognition: Papcun, Kreiman, & Davis, 1989) can impact on performance. Earwitness studies similarly report complex interactions between listener performance, stimulus duration and retention intervals (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004; Yarmey & Matthys, 1992). Other studies have described the impact of listener characteristics on speaker identity perception: listeners are, for example, more successful at recognising and learning vocal identities when exposed to speech samples produced in a language highly familiar to

them (Perrachione, Del Tufo, & Gabrieli, 2011; Perrachione, Pierrehumbert, & Wong, 2009; Zarate, Tian, Woods, & Poeppel, 2015), even when having only been passively exposed to the language (without speaking or understanding it: Orena, Theodore, & Polka, 2015). In a recent study, Lavan, Scott, and McGettigan (2016a) have shown evidence for vocalisation-specific effects during identity processing: performance on a speaker discrimination task was impaired for both familiar and unfamiliar listeners for spontaneous laughter (produced in response to genuine amusement) compared to volitional laughter (produced in the absence of genuine amusement). The authors speculate that this effect could either be grounded in the production or the perception of these vocal signals, or some combination of the two.

Spontaneous vocal signals have been shown to differ from volitional vocal signals, both in how they are produced and perceived: Distinct neural systems have been proposed to underpin the control of volitional and spontaneous laughter, respectively (Ackermann, Hage, & Ziegler, 2014; Wild, Rodden, Grodd, & Ruch, 2003). Spontaneous laughter is thought to be produced under reduced volitional control and is considered to be phylogenetically homologous with that shown in other primate species (Davila-Ross, Owren, & Zimmermann, 2009), while volitional laughter is produced under full volitional control to flexibly modulate the vocal output – a skill particularly pronounced in human vocal production compared to other primates (Pisanski, Cartei, McGettigan, Raine, & Reby, 2016). In terms of the physiological production mechanisms, Ruch and Ekman (2001) further describe

spontaneous laughter as an inarticulate vocalisation, with air being forced out of the lungs in a largely uncontrolled way and only few supralaryngeal modulations (through the movement of articulators) being apparent. During volitional laughter, we may approximate these spontaneously occurring mechanisms within controlled laughter production (cf. McKeown, Sneddon, & Curran, 2015 for a discussion of an evolutionary arms race for laughter perception and production). These differences in control and production may result in different types of information being encoded in more or less reliable ways for volitional and spontaneous laughter. Hence, our finding of impaired speaker identity discrimination in spontaneous laughs may reflect impoverished encoding of identity characteristics in the productions of these laughs, relative to volitional laughter sounds.

In perception, listeners are able to readily discriminate between spontaneous and volitional laughter (Bryant & Aktipis, 2014; Lavan, Scott, & McGettigan, 2016b), with neuroimaging studies reporting sensitivity to differences in laughter authenticity even during passive listening (McGettigan et al., 2015). It has been shown that emotional content can capture a perceivers' attention (Grandjean et al., 2005; Öhman, Flykt, & Esteves, 2001; Sander et al., 2005) – in a similar vein, other studies have suggested that the processing of this salient emotional information may be prioritized over the processing of (in some contexts) minimally salient identity information (Goggin, Thompson, Strube, & Simental, 1991; see Stevenage & Neil, 2014 for a review). Such effects of attentional capture or perceptual prioritization may differentially affect volitional and spontaneous laughter due to their distinct properties. For example, only laughs that are perceived to be high in authenticity may be affected by attentional capture.

Thus, volitional and spontaneous laughter differ in various aspects of their production and perception. It is unclear whether, and to what extent, each of these properties affects speaker identity processing. Addressing this issue has important theoretical and methodological implications: If perceptual properties (i.e. the perceived authentic emotional content in laughter) have an effect, this would provide direct empirical evidence for identity and affective information interacting during voice processing - popular models of voice perception have suggested that these types of information are processed in a largely independent fashion (see Belin, Bestelmeyer, Latinus & Watson, 2011). If production mode (contrasting volitionally versus spontaneously produced laughter) has an effect, this would call for a reframing and re-evaluation of our understanding of speaker identity perception - most previous studies have solely investigated vocal identity using subsets of volitional vocalisation types (i.e. speech), while spontaneous behaviours such as laughter have largely been ignored.

In the current study, we therefore manipulated the perceived authenticity of two types of laughter - volitional and spontaneous - to test the relative impact of laughter perception and production on identity processing. We selected four sets of laughs that systematically varied in production mode and perceived authenticity: 20 volitional laughs that were low in perceived authenticity (Volitional$_{Low}$), 20 spontaneous laughs that were perceived as being high in authenticity

(Spontaneous$_{High}$) plus additional sets of volitional and spontaneous laughter that were selected to have matched authenticity in the mid range (Volitional$_{Mid}$ and Spontaneous$_{Mid}$). We presented participants with permuted pairs of these laughter sets and asked them to discriminate speaker identity from within each pair. This design allowed us to make two distinct sets of predictions for speaker discrimination performance, one modeling production mode as the driving factor (Fig. 1a) and one based on a primary role for perceived authenticity of laughter (Fig. 1b). If production mode has an effect on speaker discrimination, performance should be similar between the two conditions including volitional laughter (Volitional$_{Mid}$ and Volitional$_{Low}$), and between the two conditions including spontaneous laughter (Spontaneous$_{Mid}$ and Spontaneous$_{High}$), with an overall advantage for volitional compared with spontaneous conditions (see Lavan et al., 2016a who show an impairment of speaker discrimination in spontaneous laughter). If key perceptual features, such as perceived authenticity, affect listeners' ability to discriminate between speakers, we should observe that performance in the speaker discrimination task should decrease with increasing perceived authenticity. This would results in performance being highest for Volitional$_{Low}$, while performance for Spontaneous$_{Mid}$ and Volitional$_{Mid}$ should be similar due to their matched properties. Performance should be lowest for Spontaneous$_{High}$, since the perceived authenticity for this condition is highest.

Further conditions were included that featured mixed category pairs of vocalisations (see Methods). Here, listeners were required to discriminate speakers from pairs that included comparisons across production mode and/or across perceived authenticity categories. Based on the findings of Lavan et al. (2016a) showing detrimental effects for pairs going, for example, across vocalisation categories, we predicted that performance should be generally lower for mixed trials compared to those within production mode, or comprising sounds from matched-authenticity sets.

2. Methods

2.1. Participants

50 participants (29 female; M_{Age}: 23.85 years; SD: 4.91 years; range 18–42 years) were recruited at Royal Holloway, University of London and University College London. This sample size was deemed adequate as similar studies of this nature have reported reliable effects with smaller sample sizes (Lavan et al., 2016a; N = 23 and N = 43), and because we anticipated that a subset of participants would need to be excluded (see Design and Procedure for exclusion criteria). Participants were paid at a rate of £7.50/h. All participants reported normal or corrected-to-normal vision, and did not report any hearing difficulties. Ethical approval was obtained from the Departmental Ethics Committee at the Department of Psychology, Royal Holloway, University of London and the Institute of Cognitive Neuroscience at University College London. None of the participants was familiar with the speakers used.
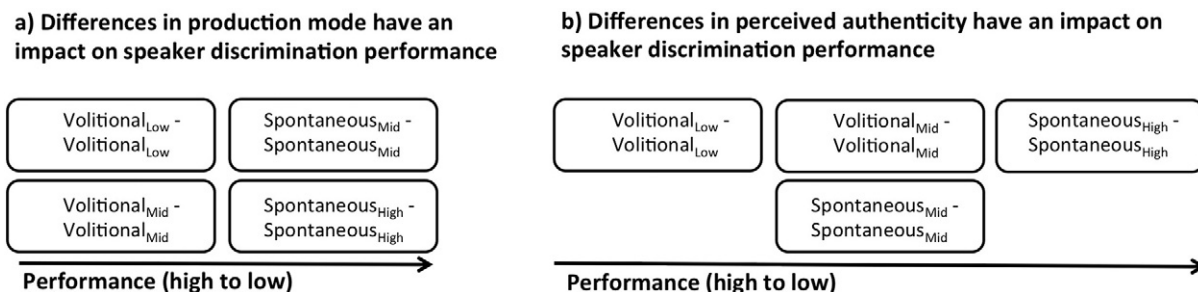

Fig. 1. Illustration of predictions: a) predicted results if speaker discrimination performance is mainly affected by production mode, b) predicted results if speaker discrimination performance is mainly affected by perceived authenticity.