Impoverished encoding of speaker identity in spontaneous laughter

Nadine Lavan, Bethanie Short, Amy Wilding & Carolyn McGettigan

[1] Department of Psychology, Brunel University, London, UK

[2] Department of Psychology, Royal Holloway, University of London, Egham, UK

[3] Institute of Cognitive Neuroscience, University College London, London, UK

Word count abstract: 193

Total word count (excl. abstract and references): 4913

Abstract

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 the 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.

Keywords: Speaker identity; voice; volitional vocalisation; laughter; speaker discrimination

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 and 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 recognizing and learning vocal identities when exposed to speech samples produced in a language highly familiar to them (Perrachione, Pierrehumbert & Wong, 2009; Perrachione, del Tufo & Gabrieli, 2011; Zarate, Xian, 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 vocalization-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, 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 & Zimmerman, 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, Walsh, Jessop, Agnew, Sauter, Warren & Scott, 2015). It has been shown that emotional content can capture a perceivers' attention (Öhman, Flykt, & Esteves, 2001; Grandjean et al., 2005, 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 vocalization types (i.e. speech), while spontaneous behaviors 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 processes 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 (Figure 1a) and one based on a primary role for perceived authenticity of laughter (Figure 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., 2016 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 $\text{Volitional}_{\text{Low}}$, while performance for $\text{Spontaneous}_{\text{Mid}}$ and $\text{Volitional}_{\text{Mid}}$ should be similar due to their matched properties. Performance should be lowest for $\text{Spontaneous}_{\text{High}}$, since the perceived authenticity for this condition is highest.

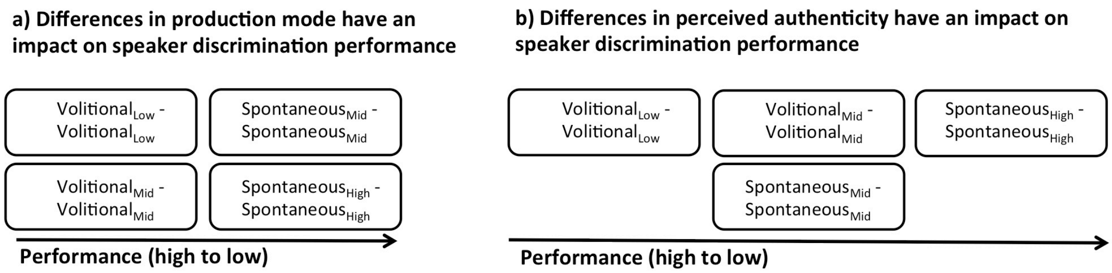

Figure 1 Illustration of predictions: a) predicted results if speaker discrimination performance is mainly affected by production mode, b) predicted results if speaker discrimination performance is mainly affected by perceived authenticity.

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 vocalization 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.

Methods

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 per hour. 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.

Materials

Spontaneous (authentic) laughter and volitional laughter from 5 speakers (all female[1], age range 29 – 34 years) were recorded in a sound-treated recording booth at Royal Holloway, University of London. For the purposes of this study, we operationalize spontaneous laughter as laughter elicited from speakers who watched or listened to amusing sound or video clips. While no detailed ratings of the speakers' emotional state were collected, speakers reported genuine feelings of amusement during and

[1] Note that the key finding that speaker discrimination performance is impaired in spontaneous laughter has previously been shown in stimulus sets including female speakers only, and including male and female speakers (see Lavan et al., 2016a).

after the recording of spontaneous laughter. We further operationalize volitional laughter as laughter produced without inducing a specific emotional state, following the instructions to produce laughter while sounding as natural as possible (see McGettigan et al., 2015; Lavan et al., 2016b for similar methods). Volitional laughs were recorded in the same session as spontaneous laughter, with volitional laughter always being recorded first to avoid carry-over effects. Recordings were obtained using a Røde condenser microphone (NT-A) with a sampling rate of 44100 Hz. The output of the microphone was fed into a PreSonus Audiobox that was connected to the USB port of the recording computer. Participants were asked to remain as still as possible during the recordings, but were seated at a distance of about 50cm from the microphone to avoid that any movement associated with intense laughter would interfere with the recordings or move the microphone. This procedure resulted in recordings per speaker and laughter type lasting several minutes (depending on the length of the videos that the speakers viewed, and the frequency of laughter events) and thus included a variable number of laughs per speaker. All perceptible laughs were extracted from the raw recordings and saved as uncompressed WAV files (min: 13, max: 52 [note: this speaker produced a large number of short laughs] per laughter type per speaker). To limit the number of laughter tokens to be included in the pilot study (see below) and to allow for a more controlled stimulus set, short (< 1.2 seconds) and very long laughs (> 3.3 seconds) were excluded. Furthermore, we excluded all recordings including background noise, clipped laughs due to excessive loudness, and recordings including breathing distortions. This pre-selection resulted in 89 volitional laughs (minimum: 9 laughs per speaker; maximum of 15) and 93 spontaneous laughs (minimum: 11 laughs per speaker; maximum of 17). For sample stimuli, please see the supplementary materials.

Perceptual features

The preselected laughs were included in a pilot study to measure the perceptual properties of the stimuli. 12 participants provided ratings of perceived arousal ("*How aroused is the person producing the vocalization?*", with 1 denoting "*the person is feeling very sleepy and drowsy*" and 7 denoting "*the person is feeling very alert and energetic*", see Russell [1980]) and perceived authenticity ("*How authentic is the vocalization?*", with 1 denoting "*not authentic at all*", *that is laugh is very posed or fake* and 7 denoting "*very authentic*", *the laughter is genuine*). At this stage, one laugher was excluded because the majority of her laughs were unvoiced (see Bachorowski & Owren, 2001)., while the remaining speakers had produced voiced laughter. Based on this pilot study, 20 volitional and 20 spontaneous laughs (5 items each from 4 laughers) were selected from the mid-range in perceived authenticity and arousal (average perceived authenticity ratings: $\text{Volitional}_{\text{Mid}}$ = 4.25, SD = 0.6; $\text{Spontaneous}_{\text{Mid}}$ = 4.29, SD = 0.27; $t[38] = .296$, $p = .769$, Cohen's $d = .088$[2]; average perceived arousal ratings: $\text{Volitional}_{\text{Mid}}$ = 4.6, SD = 0.53; $\text{Spontaneous}_{\text{Mid}}$ = 4.56, SD = 0.42, $t[38] = .259$, $p = .797$, Cohen's $d = .086$). Two additional sets of laughs were selected: one set of volitional laughs with lower perceived authenticity, and one set of spontaneous laughs with higher perceived authenticity, compared with the matched sets ($\text{Volitional}_{\text{Low}}$ = 3.07, SD = 0.69; $\text{Volitional}_{\text{Mid}}$ = 4.25, SD = 0.6, $t[38]$= 5.826, $p < .001$, Cohen's $d = 1.872$; $\text{Spontaneous}_{\text{Mid}}$ = 4.29, SD = 0.27; $\text{Spontaneous}_{\text{High}}$

[2] Cohen's *d* for independent samples was calculated using the formula (2) referred to in Lakens (2013). For dependent samples, formula (7) in Lakens (2013) was used.

= 5.6, SD = 0.36: t[38]= 13.416, p < .001, Cohen's d = 4.224). A similar pattern was apparent for perceived arousal in spontaneous laughter ($Spontaneous_{Mid}$ = 4.56, SD = 0.42; $Spontaneous_{High}$ = 5.5, SD = 0.39: t[38]= 7.369, p < .001, Cohen's d = 2.466), while perceived arousal was not significantly different for $Volitional_{Low}$ and $Volitional_{Mid}$ ($Volitional_{Low}$ = 4.35, SD = 0.55; $Volitional_{Mid}$ = 4.6, SD = 0.52, t[38]= 1.083, p = .286, Cohen's d = .479).

Acoustic features

8 key acoustic features (total duration, F0 Mean, F0 SD, Spectral Centre of Gravity, Percentage of Unvoiced segments and Harmonics-to-Noise Ratio [HNR], see Lavan, Scott & McGettigan, 2016) were extracted. An overview of the acoustic properties of the four resulting stimulus sets is given in Table 1. ANOVAs were run to determine differences in acoustic features for selected contrasts between laughter sets ($Spontaneous_{Mid}$ versus $Spontaneous_{Low}$; $Volitional_{Mid}$ versus $Volitional_{Low;}$ $Spontaneous_{Mid}$versus $Volitional_{Mid}$; α = .007 Bonferroni-corrected for 7 comparisons) – each test modeled laughter sets as a fixed effect and speaker identity as a random effect. These tests revealed that the acoustic features of laughter were matched for all comparisons of interest. Speaker had no significant effect on any of the comparisons.

Table 1 Overview of mean and standard deviations of key acoustic features of the laughter sets used in the study. Total duration indexes the average duration of stimuli in seconds. Burst duration reports the average duration of voiced laughter bursts within the stimuli. F0 Mean measures the average F0 in Hertz of the laughs. F0 SD measures the F0 variability within a laugh. Spectral Centre of Gravity measures the mean height of the frequencies for each vocalization, which captures the weighting of energy in the sound across the frequency range. Percentage of Unvoiced Segments quantifies the percentage of frames lacking harmonic structure. HNR stands for harmonics-to-noise-ratio and is defined as the mean ratio of quasi-periodic to non-periodic signals across time segments.

Sound	**Spontaneous$_{High}$**		**Spontaneous$_{Mid}$**		**Volitional$_{Mid}$**		**Volitional$_{Low}$**	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Total Duration (secs)	2.4	*0.37*	1.51	*0.41*	2.23	*0.59*	1.77	0.54
Burst duration (secs)	0.14	*0.067*	0.14	*0.079*	0.13	*0.034*	0.15	0.074
F0 Mean (Hz)	533.3	*96.79*	407.41	*107.41*	401.24	*79.13*	340.83	54.36
F0 SD (Hz)	134.18	*57.56*	117.96	*86.39*	135.33	*68.47*	96.16	41.96
Spectral Center of Gravity (Hz)	1305.55	*563.97*	941.73	*270.75*	1048.89	*273.17*	1006.54	252.88
% Unvoiced Segments	74.34	*12.16*	66.09	*11.31*	70.87	*10.6*	62.83	10.15
HNR (dB)	7.94	*2.82*	7.37	*3.04*	6.21	*1.63*	7.00	1.64
Burst duration (secs)	0.15	*0.046*	0.13	*0.046*	0.13	*0.054*	0.15	0.85

Design and Procedure

Speaker discrimination task

Participants heard permutations of pairs of laughter sounds assembled from the different categories. The two sounds were presented sequentially with a silent interval of 0.7 seconds between them. We chose 8 different pairs to address our predictions: 4 matched pairs, constructed to include laughter tokens that are within production mode and within perceived authenticity categories (Volitional$_{Low}$-Volitional$_{Low}$, Volitional$_{Mid}$-Volitional$_{Mid}$, Spontaneous$_{High}$-Spontaneous$_{High}$, Spontaneous$_{Mid}$- Spontaneous$_{Mid}$) and 4 mixed category pairs, where pairs were mismatched across production mode and/or perceived authenticity categories ($\text{Volitional}_{\text{Mid}}$-$\text{Volitional}_{\text{Low}}$, $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{Mid}}$, $\text{Volitional}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$, $\text{Spontaneous}_{\text{High}}$-$\text{Volitional}_{\text{Low}}$). Participants were not pre-informed about the inclusion of spontaneous and volitional laughter in the tasks. There were 40 trials for each pair, with 20 trials comprising two sounds from the same speaker and 20 trials presenting two sounds from different speakers – this yielded 320 trials in total. Each participant heard a subset of all of the possible pairings of speakers, with all speakers being presented an equal number of times. The order of presentation for the two sounds within a trial was counterbalanced – for instance, for $\text{Volitional}_{\text{High}}$-$\text{Volitional}_{\text{Low}}$ trials, half began with $\text{Volitional}_{\text{Low}}$ and half began with $\text{Volitional}_{\text{High}}$. After the presentation of the sounds, participants were asked to indicate via a button press on a keyboard whether they thought the two sounds were produced by the same speaker or by two different speakers. Reaction times were recorded but due to a lack of predictions these were not analysed.

Perceptual ratings task

After the speaker discrimination task, participants provided ratings of perceived arousal and perceived authenticity, in a design identical to the pilot study. The arousal scale always preceded the authenticity scale in order to avoid influencing arousal judgements through explicit knowledge of the presence of volitional and spontaneous laughter. Participants were presented with the sounds over headphones and gave their responses by pressing a key on the keyboard. Trials were timed, giving participants 3 seconds to respond before automatically moving on to the next trial. This task was included to confirm the ratings from the pilot study and to assess how perceived authenticity on a per-participant level affects performance in a speaker identity task. For the group analyses reported below, we included only participants who did not perceive a difference in perceived authenticity between the $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ conditions – this was assessed by per-participant independent samples t-tests on trial-wise ratings. This restriction was imposed on the data so as not to bias our analyses towards the production account (see Figure 1a), ensuring that on a per-items level, each subject perceived $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ to be similarly in authenticity. Using this criterion, 13 out of 50 participants were excluded (note that results for the full data set were very similar to the restricted data set – an overview over the main analyses including all 50 participants can be found in the Supplemental Materials). In the following section, the comprehensive results of all planned analyses are reported. All conditions and measures that were included in the study are reported in this paper. The presented data are all new data and have not been published before.

Results

Perceptual ratings task

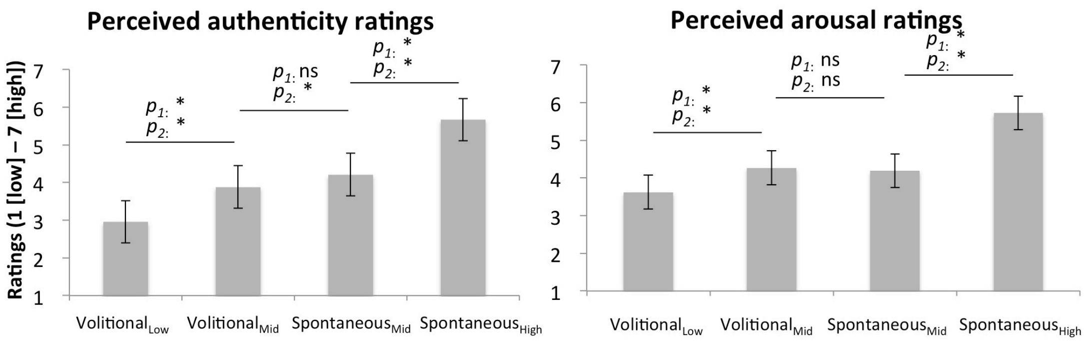

Figure 2 Results of the participants' ratings of perceived authenticity (left panel) and perceived arousal (right panel). Error bars are +/- 1 SEM. Asterisks highlight significant differences after correcting for three comparisons (Bonferroni; α = .017; p_1 = results from per-item analyses, p_2 = results from per participant analyses).

Based on a per-participant items analysis, we included only those 37 participants for whom there was no significant difference in perceived authenticity ratings between the $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ sets. $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ were thus matched within-participant for perceived authenticity. In the following section, we report the results of per-item (t_1, independent samples t-tests) as well as per-participant (t_2, dependent samples t-tests) analyses of the results of the perceptual ratings task. Results are corrected for three comparisons (Bonferroni, α = .017). Significant differences in authenticity ratings emerged for contrasts $\text{Volitional}_{\text{Low}}$ and $\text{Volitional}_{\text{Mid}}$ (t_1[72] = 5.573, p_1 < .001, Cohen's d = 1.296 CI [.595, 1.257]; t_2[36] = 11.172, p_2 < .001, Cohen's d = 1.837, CI [.758, 1.09]). Similarly, significant differences in authenticity ratings emerged for contrasts $\text{Spontaneous}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{High}}$ (t_1[72] = 10.903, p_1 < .001, Cohen's d = 2.535, CI [1.922, 1.736]; t_2[36] = 19.418, p_2 < .001, Cohen's d = 3.192, CI [1.307, 1.611]). For the contrast, $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$, no significant effect emerged for the per-item analysis (t_1[72] = 2.185, p_1 = .032, Cohen's d = .508, CI [.029, .328]), while there was a significant effect in the per-participant analysis (t_2[36] = 4.915, p_2 < .001, Cohen's d = .808, CI [.192, .463]). A similar analysis was conducted for arousal ratings: Ratings were significantly different from each other for $\text{Volitional}_{\text{Low}}$ and $\text{Volitional}_{\text{Mid}}$ as well as $\text{Spontaneous}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{High}}$ (all ps < .001, Cohen's ds > 1.59), but not for $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ (t_1[72]= .905, p_1 = .570, Cohen's d = .133, CI [-.197, .355]; t_2[36]= 1.124, p_2 = .268, Cohen's d = .185, CI [-.063, .221]; see Figure 2). The ratings largely reflect the pilot ratings used to select the stimuli for the different laughter conditions and thus validate the experimental manipulation of perceived authenticity. While for this set of ratings, a significant difference between $\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$ was found in a per-participant analysis, the magnitude of this effect is smaller compared to the magnitude of the effect for the remaining two comparisons of interest and no such effect was found in a per-item analysis (see Figure 2, left panel).

Speaker discrimination task

D' scores were calculated from the raw responses. Hit and False Alarm rates of 1 and zero were adjusted using the formula ($[n - 0.5] \div n$; n = number of trials per pair; see Stanislaw & Todorov, 1999) for all analyses. After this adjustment, the highest possible *d'* value was 3.72. We entered these data into a one-way repeated measures ANOVA with pair as a factor with 8 levels. There was a main effect of pair ($F[7, 252] = 19.684$, $p < .001$, $\eta_p^2 = .354$).

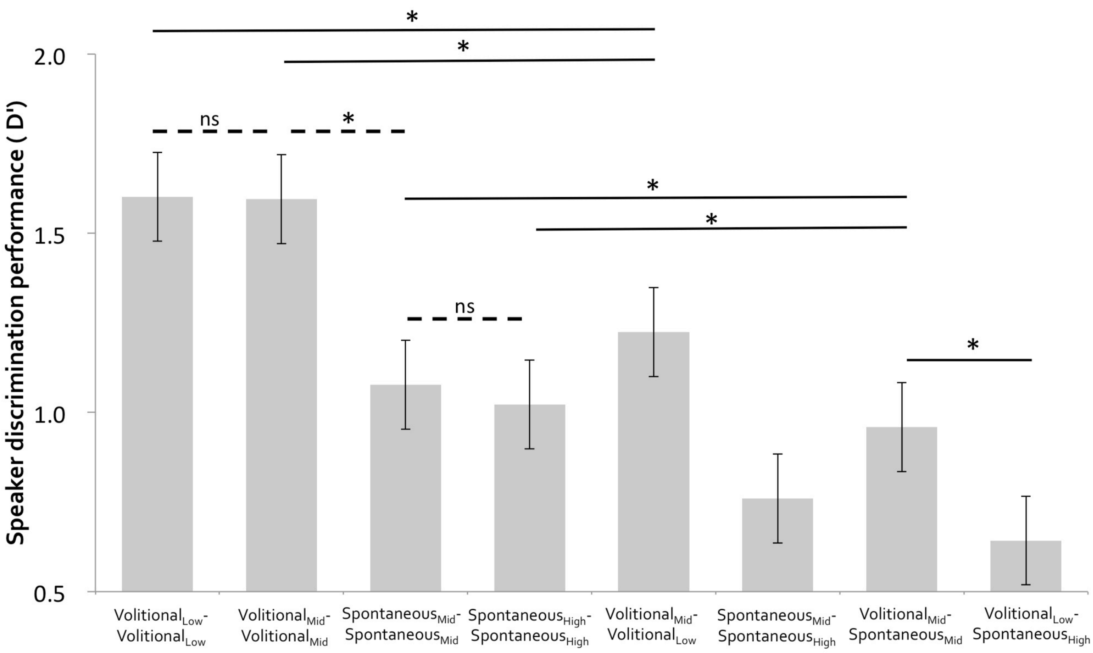

Figure 3 Results of the speaker discrimination task. Error bars show +/- 1 SEM. Asterisks highlight significant differences after correcting (Bonferroni correction) for multiple comparisons (3 planned contrasts exploring effects production versus perception [$\alpha = .017$], 5 planned contrasts exploring the effects of across production mode and/or across perceived authenticity category judgements [$\alpha = .01$]). Dashed lines indicate comparisons for testing the effects of perception versus production on performance, solid lines indicate comparisons assessing the effects of mixed category judgements within pairs on performance. For accuracy scores, see supplementary materials (Supplementary Figure 1). For response bias analyses (C'), please see Supplementary Figure 2)

Does perceived authenticity or production mode affect performance?

To address our competing hypotheses about perception and production effects (see Figure 1), an one-way ANOVA and three planned contrasts were performed on all within-production mode pairs (Figure 3, dashed lines; Bonferroni-corrected for 3 comparisons, $\alpha = .017$). This ANOVA showed that there was a significant effect of condition on speaker discrimination performance ($F[3, 108] = 16.241$, $p < .001$, $\eta_p^2 = .311$). The planned contrasts showed no difference in performance between $\text{Volitional}_{\text{Low}}$-$\text{Volitional}_{\text{Low}}$ and $\text{Volitional}_{\text{Mid}}$-$\text{Volitional}_{\text{Mid}}$, ($t[36] = .064$, $p = .949$, Cohen's $d = .014$, CI [-.231, .246]) nor between $\text{Spontaneous}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{High}}$ ($t[36] = .555$, $p = .582$, Cohen's $d = .091$, CI [-.142, .249]). However, the difference between $\text{Volitional}_{\text{Mid}}$-$\text{Volitional}_{\text{Mid}}$ and $\text{Spontaneous}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$ was significant ($t[36] = 4.337$, $p < .001$, Cohen's $d = .713$, CI [.276, .762]), with participants showing lower speaker discrimination accuracy for the spontaneous laughs. This analysis supports that production mode (see Figure 1a), but not perceived authenticity (or arousal), has an impact on speaker discrimination performance: performance remained the same despite differences in perceived authenticity/arousal while speaker discrimination accuracy differed across spontaneous and volitional laughter types that were matched in perceptual authenticity and arousal. For an analysis exploring the effects of acoustic properties on the speaker discrimination task, see the Supplementary Materials.

Effects of judgements across production mode or perceived authenticity within pairs

To explore whether judgements across perceived authenticity category and/or across production mode affect performance in a detrimental way (see Lavan et al., 2016a), five planned pairwise contrasts were performed (see Figure 3, solid lines, Bonferroni-corrected for 5 comparisons, α = .01): These planned contrasts confirmed that discrimination performance on pairs comprising a combination of high/low versus mid authenticity items (i.e. $\text{Volitional}_{\text{Mid}}$-$\text{Volitional}_{\text{Low}}$ and $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{Mid}}$) was lower than on the corresponding matched category pairs (*p*s ≤ .009, with the exception of $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{High}}$ versus $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{Mid}}$: $t[36]= 2.451$, $p = .019$, Cohen's $d = .403$, CI [.0453, .481]). Similarly, performance for $\text{Volitional}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$ was better $\text{Volitional}_{\text{Low}}$-$\text{Spontaneous}_{\text{High}}$ ($t[36] = 2.71$, $p = .01$, Cohen's $d = .446$, CI [.08, .554]). Thus when listening to sounds differing in laughter type *and* perceived authenticity, we find that performance is lower than when making judgements across laughter types where perceived authenticity is matched. This is a confirmation that increasing dissimilarity within a pair (be that in production mode, perceptual qualities or even acoustic features, which is not tested here) has additive detrimental effects on speaker discrimination performance.

Discussion

The current study set out to explain the compelling observation that speaker identity processing is significantly impaired for spontaneous laughter vocalizations. Specifically, we separated the effects of laughter production mode (volitional laughter versus spontaneous laughter[3]) and perceived authenticity on listeners' accuracy when performing a speaker discrimination task from vocal signals. We were thus able to assess whether speaker identity is less successfully encoded in spontaneous laughter (compared to volitional laughter) or whether speaker discrimination is rather affected by attentional capture based on salient acoustic features. The current results suggest that production mode has a consistent effect on speaker discrimination performance, while perceived authenticity has little to no effect on listeners' ability to discriminate between identities: performance was comparable for $\text{Volitional}_{\text{Mid}}$-$\text{Volitional}_{\text{Mid}}$ versus. $\text{Volitional}_{\text{Low}}$-$\text{Volitional}_{\text{Low}}$ as well as for $\text{Spontaneous}_{\text{High}}$-$\text{Spontaneous}_{\text{High}}$ versus. $\text{Spontaneous}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$. Crucially, however, there was a significant difference in speaker discrimination accuracy between $\text{Spontaneous}_{\text{Mid}}$-$\text{Spontaneous}_{\text{Mid}}$ and $\text{Volitional}_{\text{MId}}$-$\text{Volitional}_{\text{Mid}}$, despite being matched for perceived authenticity on a per-subject level.

Discrimination of speaker identity within sound pairs that crossed categories of perceived authenticity (low/mid/high) was generally lower than for matched conditions (i.e. within-production-mode pairs with similar levels of perceived authenticity). We suggest, however, that this may not be an effect of perceived authenticity *per se*: in this study, differences in perceived authenticity were

---

[3] We again note that no explicit ratings of subjective amusement of the speakers are available for the stimulus. Speakers did, however, report genuine feelings of amusement while recording the spontaneous laughter (and no amusement while recording the volitional laughter); see Methods for details.

additionally associated with differences in key acoustic properties, such as F0 mean (see Table 1). For the mixed categories, listeners were thus required to compare two laughs that were likely to be less acoustically similar than for matched category pairs where items are selected from within the same set. In line with this argument, Lavan et al. (2016a) have previously shown that unfamiliar listeners struggle to generalize identity-related information across pairs of vocalisations that are dissimilar to each other (e.g. series of vowels compared with laughter).

The results of current study thus suggest that information about identity is less successfully encoded in spontaneous laughter vocalisations. There is a body of literature that argues that the production mechanisms underlying spontaneous vocalisations differ from those generating volitional vocalisations: spontaneous vocalisations are homologous to those observed in other primate species, and can be directly linked and likened to animal vocalisations based on their acoustic properties (Davila-Ross et al., 2009). This link is also reflected on a neural level, with a phylogenetically homologous pathway being thought to underpin spontaneous vocal production, while the pathway involved in volitional vocal production shows features that are unique to humans (Ackermann et al., 2014). Related to this, Bryant and Aktipis (2014) show that when slowing down examples of spontaneous laughter, human listeners could not distinguish these laughs from slowed down animal calls, while slowed down volitional laughs were still identifiable as human vocalisations.

Sidtis and Kreiman (2012) propose that our ability to process voice identity is prodigious compared to other species and, much like human language or speech processing, unique. While other species can to some extent distinguish voices as those of their kin or familiar individuals, or discriminate members of their own species from unknown others (Kreiman & Sidtis, 2011), there is no known upper estimate on how many (familiar) voices a human can recognise. In humans, performance for discriminating speakers based on short samples of speech has consistently been shown to be very high (Reich & Duke, 1979; Van Lancker & Kreiman, 1987; Wester, 2012). Based on our current finding, we suggest that the potentially unique ability to process identity-related information from vocal signals in humans may be a results of the ability of our species to volitionally control and modulate voice production: Through volitionally modulating and controlling vocalisations, humans may be able to produce more individuated vocal signals, encoding a wealth of identity-specific cues that may not be encoded during relatively stereotyped spontaneous vocal behaviours. Individual-specific volitional vocalisations may thus allow listeners to more easily differentiate between people based on vocal signals. In this framework, spontaneous vocalisations can thus be considered as representative of an older stage of evolution that still efficiently conveys, for example, emotional content, while volitional vocal production may have evolved to more effectively encode cues to identity, for example in adapting to larger group sizes (alongside its posited role in the evolution of spoken language; Pisanski et al., 2015). It remains to be determined whether this possible enhancement of identity encoding in vocal signals is an adaptive process in response to evolutionary pressures (e.g. distinguishing individuals despite larger social group sizes, cf Pollard & Blumstein, 2011; see also Dunbar, 1993) or merely a by-product of the evolution of speech (e.g. through the increased complexity of speech signals offering more degrees of freedom for idiosyncratic and thus diagnostic vocal production behaviours).

Our study thus adds to a growing literature on laughter as a rich signal that can be used by listeners - to varying extents - to make judgements about people, from their social relationships to identity perception (Bryant et al., 2016; Lavan et al., 2016a; Scott, Lavan, Chen & McGettigan, 2016). Our paper furthermore raises key questions for our understanding of person perception from the voice (which has hitherto almost exclusively investigated perception of volitional behaviours, such as speech) and its possible evolutionary origins. It should be noted that the current study was focused on two broad types of laughter. Laughter is, of course, only one example of how human vocalisations can vary in their production mechanisms, as well as their physical and perceptual properties. Future work should expand on these findings and determine if there is indeed a systematic advantage for all volitionally produced vocal signals or whether this is specific to laughter. Generally, our findings call for a broader and more nuanced account of identity processing that includes the full range of vocal signals frequently used in human communication, and that recognizes which features of a vocalization help or hinder the successful encoding and extraction of identity-related information.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(06), 529-546.

Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, *12*(3), 252-257.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America*, *40*(6), 1441-1449.

Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, *35*(4), 327-335.

Davila-Ross, M., Owren, M. J., & Zimmermann, E. (2010). The evolution of laughter in great apes and humans. *Communicative & Integrative Biology*, *3*(2), 191-194.

Dunbar, R. I. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences*, *16*(04), 681-694.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448-458.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature neuroscience*, *8*(2), 145.

Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. P. A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology*, *18*(3), 327-336

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*.

Lavan, N., Scott, S. K., & McGettigan, C. (2016a). Impaired generalization of speaker identity in the perception of familiar and unfamiliar voices. *Journal of Experimental Psychology: General*, *145*(12), 1604.

Lavan, N., Scott, S. K., & McGettigan, C. (2016b). Laugh like you mean it: authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, *40*(2), 133-149.

Mathias, S. R., & von Kriegstein, K. (2014). How do we recognise who is speaking. *Frontiers in Biosciences* (*6)*, 92-109.

McKeown, G., Sneddon, I., & Curran, W. (2015). Gender differences in the perceptions of genuine and simulated laughter and amused facial expressions. *Emotion Review*, *7*(1), 30-38.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalising and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, *25,* 246-257.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466-478.

Orena, A. J., Theodore, R. M., & Polka, L. (2015). Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition*, *143*, 36-40.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, *85*(2), 913-925.

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, *333*(6042), 595-595.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. (2009). Differential neural contributions to native-and foreign-language talker identification. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1950.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: a window into the origins of human vocal control? *Trends in Cognitive Sciences*, *20*(4), 304-318.

Pollard, K. A., & Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, *21*(5), 413-417.

Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *The Journal of the Acoustical Society of America*, *66*(4), 1023-1028.

Ruch, W., & Ekman, P. (2001). The expressive pattern of laughter. *Emotion, qualia, and consciousness*, 426-443.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39, 1161–1178.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, *28*(4), 848-858.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing Famous Voices Influence of Stimulus Duration and Different Types of Retrieval Cues. *Journal of Speech, Language, and Hearing Research*, *40*(2), 453-463.

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in cognitive sciences*, *18*(12), 618-620.

Sidtis, D., & Kreiman, J. (2012). In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. *Integrative Psychological and Behavioral Science*, *46*(2), 146-159.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149.

Stevenage, S., & Neil, G. (2014). Hearing faces and seeing voices: The integration and interaction of face and voice processing. *Psychologica Belgica*, *54*(3), 266-281.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, *25*(5), 829-834.

Wester, M. (2012). Talker discrimination across languages. *Speech Communication*, *54*(6), 781-790.

Wild, B., Rodden, F. A., Grodd, W., & Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, *126*(10), 2121-2138.

Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, *6*(5), 367-377.

Zarate, J. M., Tian, X., Woods, K. J., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific reports*, *5*.