

**Increased discriminability of authenticity from multimodal laughter is driven by
auditory information**

Nadine Lavan¹, Carolyn McGettigan^{1,2}

¹Department of Psychology, Royal Holloway, University of London, Egham, UK

²Institute of Cognitive Neuroscience, University College London, London, UK

Word count: 4729 words (excl. abstract and references)

Word count abstract: 148 words

Correspondence to: Nadine Lavan, Department of Psychology, Royal Holloway,
University of London, Egham Hill, Egham TW20 0EX, UK. Email:
Nadine.Lavan.2013@rhul.ac.uk.

Acknowledgements: We thank Cèsar F Lima, M.R. Manners and two anonymous
reviewers for their comments on this manuscript.

Abstract

We present an investigation of the perception of authenticity in audiovisual laughter, in which we contrast spontaneous and volitional samples and examine the contributions of unimodal affective information to multimodal percepts. In a pilot study, we demonstrate that listeners perceive spontaneous laughs as more authentic than volitional ones, both in unimodal (audio-only, visual-only) and multimodal contexts (audiovisual). In the main experiment, we show that the discriminability of volitional and spontaneous laughter is enhanced for multimodal laughter. Analyses of relationships between affective ratings and the perception of authenticity show that, while both unimodal percepts significantly predict evaluations of audiovisual laughter, it is auditory affective cues that have the greater influence on multimodal percepts. We discuss differences and potential mismatches in emotion signaling through voices and faces, in the context of spontaneous and volitional behavior, and highlight issues that should be addressed in future studies of dynamic multimodal emotion processing.

Keywords: laughter; authenticity; dynamic; audiovisual; perception

Introduction

When decoding emotional signals in everyday life, we are usually provided with emotional cues from more than one modality. A hierarchical model of emotion processing proposes that emotional information from auditory and visual channels is extracted and integrated into a unified multimodal percept, which is then explicitly evaluated (Brück, Kreifelts & Wildgruber 2011). Studies have shown that congruent emotional information from multiple modalities improves our decoding accuracy compared to unimodal emotional signals (Burns & Beier 1973; Paulmann & Pell 2011).

While one modality may suffice to decode emotional signals, increases in performance in the presence of a second modality can be observed due to at least partially independent affective information being available from each modality and also due to redundant but consequently more robust affective information from the two channels (Ay, Flack & Krakauer 2007; Campanella & Belin 2007). Studies exploring relationships between audio-only, visual-only, and audiovisual emotion perception support this claim: both audio-only and visual-only have been found to be strongly correlated with AV percepts, while correlations between A and V scores appear to be weaker or absent, suggesting partially independent contributions of each modality towards the multimodal percept (e.g. Bänzinger, Mortillaro & Scherer 2012; Burns & Beier 1973). It is thought that the integration of multiple modalities is based on a weighted linear combination of cues, dependent on the relative reliability of cues in each modality (Angelaki, Gu & DeAngelis 2009). Studies exploring how socio-emotional qualities, such as attitudes, moods and signals of agreement, are processed in multimodal contexts report a greater contribution of

visual cues compared to auditory cues (Burns & Beier 1973; Mehu & van der Maaten 2014; Mehrabian & Ferris 1967).

Studies of multimodal emotion processing to date have mostly used emotion categorization accuracy as a dependent variable. Identifying specific emotion categories is, however, only one type of judgement made when processing emotional displays: we can, for example, reliably extract broad affective qualities, such as arousal and valence, and can determine whether emotional displays are authentic. Such judgements are highly relevant to accurately decoding the nuanced information encoded in emotional signals, and warrant further investigation.

Authenticity is one such aspect of emotion processing that has recently received some attention in the literature, most notably for laughter and smiles. These studies compared volitional stimuli produced under full control with spontaneous stimuli produced under reduced control. Based on acoustic differences in the auditory modality (Jürgens, Hammerschmidt & Fischer 2011; Lavan, Scott & McGettigan 2015) and morphological differences in facial expressions (e.g. Ekman & O'Sullivan, 2006), studies have shown that individuals can distinguish between spontaneous and volitional exemplars of smiles or amusement facial expressions (Krumhuber & Manstead 2009; McKeown, Sneddon & Curran 2015) and (auditory-only) laughter (Bryant & Aktipis 2013; Audibert & Cathiard 2003). Neuroimaging studies further report differences in neural responses to spontaneous versus volitional vocalizations (McGettigan et al. 2015; Drolet, Schubotz & Fischer 2013). Beyond basic categorization accuracy, it has been shown that authentic emotional displays in both the auditory and visual domain are perceived to be more extreme in valence and higher in arousal compared to their volitional counterparts (Audibert,

Aubergé & Rilliard, 2008; Lavan et al. 2015; Wiltling, Kraemer & Swerts, 2006). Thus, authenticity has wide-ranging and varied effects on the perception of emotional signals.

While we have a basic understanding of how differences in authenticity affect the perception of emotional displays in the auditory and visual modality separately, only little is known about audiovisual authenticity processing. One study has investigated authenticity perception in multimodal emotional signals. Using a discrimination task, Audibert and colleagues (2008) investigated whether differences in authenticity in emotional signals can be perceived from audio-only, visual-only, and audiovisual stimuli, and reported that participants were able to identify spontaneous displays within pairings of spontaneous and volitional stimuli, for a range of emotions. In line with studies looking at multimodal emotion categorization, detection rates for audiovisual stimuli were higher compared to audio-only and visual-only stimuli. No difference in performance was found between audio-only and visual-only stimuli. These results allow first glimpses into how authenticity of spontaneous and volitional emotion displays is processed in a multimodal context. From those data, however, it remains unclear how the different modalities relate to each other, or how they each contribute towards multimodal percepts. The authors furthermore collapse across volitional and spontaneous vocalizations, obscuring potential differences between the two.

The current study set out to explore how authenticity in dynamic volitional and spontaneous laughter is perceived in both unimodal and bimodal contexts. By using laughter, we were able to present participants with naturalistic stimuli. In everyday life, laughter permeates all social interactions; it is produced under varying

levels of volitional control, ranging from helpless, spontaneous laughter to fully volitional vocalizations (Provine, 2000; Scott, Lavan, Chen, & McGettigan, 2014). We collected ratings of authenticity for audio-only, visual-only, and audiovisual stimuli as well as ratings of basic affective qualities - arousal and valence – in order to explore their relationships with, and contribution towards, authenticity judgements. By stepping away from traditional forced-choice categorization tasks and by instead using ratings scales, we were able to collect fine-grained measures of these continuous affective and perceptual properties of laughter, and for the first time report on within-category variability for unimodal and bimodal properties of emotional vocalizations. A pilot experiment initially explored whether participants are able to perceive differences in authenticity in laughter from dynamic audio-only, visual-only and audiovisual signals, all created from the same instances of laughter. In the main experiment, we investigated how the perception of authenticity in unimodal signals relates to the perception of audiovisual signals, focusing on the relationships between ratings of unimodal and bimodal signals.

Methods

Stimulus recording

6 volunteers (all female, $Age_{mean} = 27.5$ years; $SD = 3.78$ years; $Age_{range} = 23-33$ years) were recorded in a quiet room with a digital camcorder (Camera HC-X920) and microphone (Røde NT-A) in front of a white background. All participants wore a plain grey cap to avoid hair obscuring the face, and a black gown to further standardize the recordings. Volunteers recorded spontaneous and volitional laughter. For the elicitation of spontaneous laughter, volunteers watched video clips

that they found amusing. These video clips were presented on a laptop screen that was positioned to the right side of the camera; sound was presented via headphones. For volitional laughter, participants were instructed to produce laughter that sounded as authentic as possible, while watching videos of neutral emotional quality (short videos describing how to use statistical analysis software). After each video clip, volunteers indicated how amused they were, how authentic the laughter they produced was, and how much control they had over the production of the vocalization, on a scale from 1 (*not amused/authentic/controlled*) to 7 (*very amused/authentic/controlled*). These ratings confirm that the participants experienced more amusement (spontaneous: 5.94, volitional: 1.33) and less volitional control over production of spontaneous laughter (spontaneous: 2.56, volitional: 6.5) and thought their spontaneous laughter to be more authentic compared to their volitional laughter (spontaneous: 6.5, volitional: 1.33).

To create the stimuli, discrete laughs were identified and extracted from the raw recordings. The videos were edited to be black and white, were cropped to show the face and neck of each laugher, and based on the first frame of each individual clip, centered and aligned (aligning nose, eyes, and mouth as landmarks) using Camtasia (<https://www.techsmith.com/camtasia.html>, version 2). Stimuli where the eye gaze deviated from fixating on the screen of the laptop were excluded. The volume of the audio track was normalized for root-mean-square amplitude across all clips using PRAAT (Boersma & Weenink, 2016). From the audiovisual stimuli, visual-only and audio-only stimuli were created by either extracting the audio track or muting it in the video file. Example stimuli can be seen in Figure 1.

INSERT FIGURE 1 HERE

Pilot study

Laughter is an emotional display that affects the whole body (Ruch & Ekman, 2001; Niewiadomski, Mancini, Ding, Pelachaud, & Volpe, 2014). Here, a pilot study was conducted in order to verify that differences in authenticity and other perceptual qualities in our stimulus set can indeed be determined from audio-only, audiovisual, and visual-only stimuli derived from the same instance of laughter. In line with previous research (Audibert et al, 2008; Bryant & Aktipis, 2013; McKeown, Sneddon & Curran, 2015), we predicted that participants would be able to perceive differences in authenticity between volitional and spontaneous laughter from all presented modalities.

All extracted laughter clips between 1.2 and 3.2 seconds were used (cf. McGettigan et al. 2015; Lavan et al. 2015). After this initial preselection, a set of 50 volitional and 60 spontaneous laughs remained that were taken forward as stimuli in a pilot study. Testing sessions lasted around 1h. 16 participants ($Age_{mean} = 19.4$ years, $SD = 1.3$ years; $Age_{range} = 18-21$ years, 13 female), none of whom took part in the main experiment, were recruited from the Department of Psychology at Royal Holloway, University of London and received course credit for their participation. Participants were seated in front of a computer screen on which the stimuli were presented, with sounds being played at a comfortable volume via headphones (Sennheiser HD 201), using MATLAB (Mathworks, Inc., Natick, MA, v 2014a) with the Psychophysics Toolbox extension (<http://psychtoolbox.org/>). Participants were asked to rate audio-

only, visual-only, and audiovisual versions of the stimuli on 7-point Likert scales: ratings of 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"*), valence (*"How positive or negative is the person producing this vocalization feeling?"*, with 1 denoting *"very negative"* and 7 denoting *"very positive"*) and authenticity (*"How authentic is each laugh??"*, with 1 denoting *"not v authentic at all"* and 7 denoting *"very authentic"*). All stimuli were presented for each scale. Participants thus heard every stimulus three times throughout the experiment. Participants were not made aware that volitional and spontaneous laughter was included in the study. All trials were timed, giving participants 2.5 seconds to make a response via a key press before moving on to the next trial. Missed responses were recorded as such and excluded from the analysis. The order of presentation of the arousal and valence scale was also randomized. The authenticity scale was also presented last to avoid that the explicit knowledge of volitional and spontaneous laughter being present in the stimulus sets would have any effects of ratings of arousal and valence. Within each scale, the order of the stimuli was fully randomized. The randomization of stimulus order within each rating scale as well as the randomization of valence and arousal scales ensured that no systematic effects of stimulus presentation or scale order were possible.

Inter-rater reliability was very high on each scale, calculated across all modalities (Cronbach's $\alpha = 0.92 - 0.93$). As predicted, paired sample t-tests of average ratings per participant showed that, for all modalities, volitional laughs were rated as significantly less authentic than spontaneous ones. Further, spontaneous laughter

was higher in arousal and perceived as more positive compared to volitional laughter (see Table 1).

INSERT TABLE 1 HERE

Main experiment

Laughter has been described as a multimodal behavior, expressed in the voice, on the face and in the movements of the body (Niewiadomski, Mancini, Ding, Pelachaud, & Volpe, 2014; Ruch & Ekman, 2001). By investigating laughter as a multimodal signal, our pilot study showed that differences in authenticity cannot only be perceived from the auditory signal but also from the coincident dynamic visual-only and audiovisual cues, derived from the same instance of laughter behaviour. This extends findings of studies exploring the unimodal perception of spontaneous and volitional laughter, which have reported the successful discrimination of spontaneous and volitional laughter from visual signals (McKeown et al., 2015) and auditory-only signals (Bryant & Aktipis, 2014), and provides the basis for our main study investigating the relationships between visual and auditory modalities, and their influence on the perception of audiovisual laughter samples.

Based on the ratings collected in the pilot study, we selected a test stimulus set comprising clear exemplars of volitional and spontaneous laughter – the motivation for this selection step was to reduce noise in the data set that might be introduced by the presence of poor exemplars of these two laughter types. This set of stimuli was then employed in a perceptual study testing three main predictions. Building on the bimodal emotion recognition literature (e.g. Paulmann & Pell, 2011), we predicted that modality would have an effect on authenticity ratings, with

audiovisual ratings being rated as different from audio-only and visual-only ratings. Further, we predicted that audio-only and visual-only signals would both make significant contributions to shaping the percept of authenticity in the bimodal audiovisual stimuli. Finally, given the literature showing differential processing of volitional and spontaneous vocalizations and evidence in studies for emotion-specific profiles of responses to audio-only, visual-only, and audiovisual emotional signals (Bänzinger, Mortillaro & Scherer, 2012; Paulmann & Pell, 2011), we expected the effects of modality to differ dependent on laughter type.

Methods

Participants

44 healthy participants ($Age_{mean} = 19.2$ years, $SD = 1.4$ years; $Age_{range} = 19-21$ years, 36 female) took part in the main experiment. They were recruited at Royal Holloway, University of London and received course credit for their participation. All participants had 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.

Materials

The stimulus set was selected using authenticity ratings from the pilot study for the audiovisual stimuli, and included 24 stimuli per laughter type, with stimuli being balanced across different laughers (spontaneous laughter: 4 tokens per laugher; volitional laughter: 4 stimuli for 4 laughers, 2 and 6 stimuli respectively for the two

other laughers). Stimuli were matched for duration (spontaneous: 2.47 secs; volitional: 2.42 secs, $t[46]=.331$; $p = .742$). All selected spontaneous stimuli had an average authenticity rating above 4 and all volitional stimuli were rated below 4, ensuring clear differences in authenticity for the two types of laughter (mean ratings spontaneous laughter: 5.11; mean ratings volitional laughter: 2.64; $t[46]=12.174$; $p < .001$). Spontaneous laughter clips were also significantly higher in arousal (spontaneous: 5.16; volitional 3.37; $t[46]=9.642$; $p < .001$) and valence (spontaneous: 5.16; volitional: 3.23; $t[46]=11.894$; $p < .001$).

Procedure

Participants were presented with 144 stimuli in total (24 individual laughter stimuli x 2 types of laughter x 3 modalities) in fully randomized order. Given the number of stimuli, there was a low probability that a bimodal stimulus would be presented immediately before or after its audio-only or visual-only counterpart (and thus was not controlled for in the randomization). In a procedure identical to that used in the pilot study, participants rated arousal, valence, and authenticity for each stimulus. Testing sessions lasted up to one hour. Data for the authenticity scale from three participants was lost due to technical problems. Two data sets for the authenticity scale only were lost due to experimenter error.

Results

Difference in perceived authenticity between modalities and laughter types

To investigate how the perceived authenticity of laughter varies as a function of modality (audio-only, visual-only, audiovisual) and laughter type (spontaneous,

volitional), mean ratings per subject were computed and entered into a 2 (laughter type) x 3 (modality) repeated-measures ANOVA. Nine post-hoc paired samples t-tests were further performed to explore the significant main effects and interactions per rating scale, using a Bonferroni-corrected significance level of $p = .006$.

There were significant main effects for modality and laughter type, as well as a significant interaction (main effect of modality, $F[2,82] = 11.202, p < .001, \eta^2 = .012$; main effect of laughter type $F[1,41] = 467.127, p < .001, \eta^2 = .825$; interaction modality x laughter type, $F[2, 82] = 27.954, p < .001, \eta^2 = .027$).

The results of the post-hoc t-tests are shown in Figure 2. They confirm that the stimulus preselection indeed created two groups of laughter differing in perceived authenticity. They furthermore reveal a complex pattern of differences in authenticity ratings between modalities, with audiovisual stimuli being rated as more extreme (i.e. higher for spontaneous, and lower for volitional) compared to their unimodal counterparts mainly driving the interaction. Numerically, visual laughter was rated as more authentic compared with auditory laughter across both volitional and spontaneous laughter. To directly investigate whether the numerical trend between audio-only and visual-only ratings reflects a statistically significant interaction, we ran a repeated-measures 2 (laughter type) x 2 (modality) ANOVA, including audio-only and visual-only ratings only. The interaction was not significant ($F[1,13] = .068, p = .707, \eta_p^2 = .003$).

To formally assess whether bimodal percepts aid discriminability of volitional and spontaneous laughter, we computed within-subject difference scores of authenticity ratings (spontaneous – volitional, per modality). Three paired t-tests (one for each modality) on these scores confirmed that the perceived differences in

authenticity between volitional and spontaneous laughter were significantly bigger for audiovisual stimuli than for audio-only and visual-only stimuli (audiovisual – audio-only: $t[41]=6.121$; $p \leq .001$, Cohen's $d = .82$; audiovisual – visual-only: $t[41]=7.473$; $p \leq .001$, Cohen's $d = .96$). Differences between audio-only and visual-only ratings were similar ($t[41]=.741$; $p = .463$, Cohen's $d = .11$), indicating increased discriminability of authenticity for bimodal laughter compared to unimodal stimuli.

INSERT FIGURE 2 HERE

Contributions of auditory and visual channels towards the perception of authenticity in bimodal stimuli

To explore whether and how affective ratings for unimodal stimuli predict the bimodal ratings of authenticity for spontaneous and volitional laughter in a per item analysis, six multiple regression analyses were performed, with audiovisual ratings of authenticity as the dependent variable, and audio-only and visual-only ratings of authenticity, arousal, and valence respectively as predictors, for spontaneous and volitional laughter separately. Regressions were performed using the enter method in SPSS as we had no prediction regarding the nature or magnitude of the variance explained by auditory and visual channels. All six regression models were highly significant, showing that unimodal ratings explained a large proportion of the variance in audiovisual ratings (*adjusted* $R^2 = .60 - .90$, all $ps < .001$). For arousal and authenticity, both audio-only and visual-only ratings significantly predicted audiovisual ratings of authenticity (audio-only laughter: $\beta s = .77 - .90$, all $ps < .001$; visual-only laughter: $\beta s = .30 - .36$, all $ps < .01$). For valence, only audio-only ratings

significantly predicted audiovisual authenticity ratings (audio-only laughter: β s = .77 - .90, all $ps < .001$; visual-only laughter: β s = .12 - .18, all $ps > .136$). It was striking that visual-only ratings were consistently weaker predictors than audio-only ratings across all models. No differences were found between volitional and spontaneous laughter. Overall, these findings suggest that both unimodal channels seem to be relevant for evaluating authenticity in audiovisual laughter. Thus, ambiguity in the individual channels may be resolved through perceptual integration of auditory and visual information, based on more reliable and robust information (Ay et al., 2007).

Relationships between audio-only, visual-only, and audiovisual ratings of authenticity

Pearson's correlations were computed per item for the two laughter types, to further explore how audio-only and visual-only ratings of authenticity, arousal and valence relate to audiovisual ratings of authenticity. In line with the regression analyses, correlation coefficients were systematically lower for correlations between visual-only and audiovisual compared to those between audio-only and audiovisual, indicating differential contributions towards the audiovisual percept (see Table 2).

Hotelling-Williams tests were performed to determine the significance of the differences in audio-only - audiovisual and visual-only - audiovisual correlations. These tests confirmed that audio-only ratings were more highly correlated than visual-only ratings with audiovisual ratings of authenticity, for spontaneous laughter ($ps < .021$) and volitional laughter ($ps < .025$) and for all affective ratings scales.

INSERT TABLE 2 HERE

Intriguingly, audio-only and visual-only ratings were not significantly correlated for authenticity in volitional laughter (Pearson's $r = .289 - .303$; all $ps > .17$), but were significantly correlated for authenticity in spontaneous laughter (Pearson's $r = .443$; $p = .03$), see Table 2. Further significant audio-only – visual-only correlations were found for ratings of arousal and valence for auditory-only spontaneous laughter, and for authenticity ratings for visual-only spontaneous laughter (Pearson's $r = .417 - .418$ both $ps < .043$). None of the differences in correlation strength between volitional and spontaneous laughter was, however, significant (Fisher's z-transformation, all $ps > .17$). This lack of significant audio-only – visual-only correlations for volitional laughter corresponds to Burns' and Beier's (1973) finding. This finding is suggestive of a relative independence between the visual and auditory percepts for volitional laughter. Moderate audio-only – visual-only correlations for most comparisons in spontaneous laughter suggest a more dependent way of signaling affective information in the audio-only and visual-only channels for spontaneous laughter. Correlations between audio-only and visual-only ratings were generally weaker compared to visual-only – audiovisual correlations (Hotelling-Williams tests, volitional laughter: $ps < .001$; valence $p = .077$; spontaneous laughter: authenticity, $p = .055$; arousal, $p = .029$; valence, $p = .083$) and significantly weaker when compared with audio-only – audiovisual correlations (all $ps < .001$ for both laughter types).

Discussion

In this study, we investigated how the perception of authenticity is affected by presentation modality in spontaneous and volitional laughter. We specifically explored relationships between modalities and compared the contributions of auditory and visual channels towards an audiovisual percept of authenticity based on unimodal ratings of arousal, valence, and authenticity. In line with previous research on emotional speech (Audibert et al., 2008), we find that audiovisual laughs are, in the case of spontaneous laughter, rated as more authentic, and, in the case of volitional laughter, as less authentic compared to their unimodal counterparts (cf. Burns & Beier, 1973; Paulmann & Pell, 2011). We furthermore directly show for the first time that the discriminability of authenticity in laughter is aided by bimodal presentation. This is in line with previous findings showing increases in performance or accuracy for audiovisual stimuli, compared to unimodal stimuli. The partially independent affective information encoded in unimodal channels, alongside some redundancies in the information conveyed, will lead to a more informative as well as robust affective information being available (Ay, Flack & Krakauer, 2007; Belin & Campanella, 2007).

We further assessed to what extent auditory and visual modalities contribute towards the perception of authenticity in a bimodal context: while audio-only and visual-only ratings of arousal and authenticity significantly predicted audiovisual ratings of authenticity (for valence, only auditory ratings significantly predicted audiovisual ratings of authenticity), the prediction strength was consistently higher for audio-only ratings compared to visual-only ratings. Previous studies report a relative dominance of the visual channel over the auditory channel for the extraction of socio-emotional stimulus qualities (Burns & Beier, 1973; Mehu & van

der Maaten, 2014; Mehrabian & Ferris, 1967). Our results, however, show an auditory dominance. This may be an emotion-specific or context-specific effect: given that weightings and perception of cues seem to be dependent on the perceptual features being extracted and on the stimulus itself (Ekman, Friesen, O'Sullivan & Scherer, 1980), emotion- and judgement-specific dependencies can be expected. Indeed, previous studies have already reported emotion-dependent modality effects (e.g. Bänzinger, Mortillaro & Scherer, 2012; Paulmann & Pell, 2011). Laughter can be considered to be a primarily auditory signal - we may therefore be more attuned to extracting affective information from the auditory channel despite this information also being available in the visual signal. Perceptual profiles could be different for other non-verbal emotional signals: for example, a salient visual cue to sadness intensity is the presence of tears (Provine, Krosnowski & Brocato, 2009), while authentic *auditory* crying can be emotionally ambiguous to the extent of being confusable with (authentic) laughter (Lavan, Lima, Harvey & McGettigan, 2014). This may have an impact on the weighting of cues for affective evaluations of crying. Similarly, from the perspective of the encoder, emphases on auditory and visual signals may be context-dependent. For example, in challenging listening environments, it could be expected that encoders would enhance signal clarity or intensity in the visual domain, changing weightings and informativeness of the respective channels for a perceiver (McKeown, 2015). Given such findings, future studies will need to systematically explore emotion-dependent and context-dependent effects and establish which factors underlie such visual- or auditory-dominant influences on bimodal emotion processing. Further studies could also explore whether degrading the signal in one modality will affect weighting of cues.

Against our predictions, the relationships between unimodal and bimodal percepts were not found to be different across volitional and spontaneous laughter. Some differences were apparent in the patterns of correlations, with ratings for unimodal channels being dependent or correlated for ratings of spontaneous laughter but independent and not correlated for volitional items (cf Burns & Beier, 1973). Such a pattern could arise from differential levels of control and automaticity present during the production of these two types of laughter: volitional laughter is produced under full voluntary control of the producer, while spontaneous laughter is produced under reduced volitional control, with a higher degree of automaticity. Spontaneous laughter is a behavior that affects the whole body, with body posture, facial expression, and vocal output are affected as the result of the spontaneous spasming of thoracic muscles (Ruch & Ekman, 2001). When producing spontaneous laughter, all of the processes are automatically coordinated and synchronized. To volitionally produce and, crucially, coordinate all of these hallmarks of spontaneous laughter may be challenging, leading to mismatches in the resultant affective percepts (e.g. differences in the magnitude and/or temporal synchronization of affective cues across the two modalities). For example, audio-visual – visual-only mismatches in intensity have been shown to reduce the perceived plausibility of laughter, a concept closely linked to authenticity (Niewiadomski et al., 2015). These observations of differences in audio-only – visual-only relationships across authenticity offer some insights into the perception and production of volitional and spontaneous laughter – however, we caution that, while interesting, statistically non-significant trends should not be overinterpreted without further investigation.

As a limitation for this study, it is noted that we used recordings of 6 female laughers and did not include any male laughs. Furthermore, our participant samples were biased towards female participants. While gender effects on laughter perception and production have been shown with regard to authenticity judgements (e.g. McKeown et al. 2015), most studies reporting marked gender differences in laughter concern the production and occurrence of laughter in conversation (e.g. Lampert & Ervin-Tripp, 2006; Provine, 1993). The current study does not investigate laughter production within social contexts, but perceptual properties of multimodal laughter outside social context. While gender effects may in principle be present for authenticity perception in laughter, it was not the main aim of the study to investigate those. This, however, means that a parsimonious interpretation of our study does not allow us to generalize the findings beyond female laughter being perceived by mainly female listeners¹.

Our results contribute to the growing body of literature investigating multimodal emotion processing using dynamic stimuli. When dealing with dynamic multimodal stimuli, it is crucial to understand how perceptual and physical properties of both auditory and visual signals evolve over time, and how these time courses relate to each other. To date, these relationships have barely been explored.

¹ Note that we have conducted an analyses of the authenticity ratings from our main experiment as an example to check for the presence or absence of gender effects, looking at female listeners only ($N = 36$) and comparing this data to the male listener sample ($N=8$). Regression models for authenticity confirm the results reported for all participants: auditory information is a stronger predictor of audiovisual ratings for female and male observers even when analysed separately. The finding showing increased discriminability for audiovisual stimuli holds for the female sample, when male listeners are excluded (A-AV & V-AV, $p < 0.001$, A-V, $p = .7$) while no significant effect was found for the male sample (all $ps > .09$). It should, however, be noted that this effect also did not reliably replicate (i.e. patterns of significance differed across samples) in three randomly selected samples of female listeners of the same size as the male sample (all $N = 8$). We therefore would like to suggest that, while these analyses should not be overinterpreted, there is no concrete evidence for meaningful gender effects in our study.

Linking the perception of affective qualities to physical properties (e.g. acoustic features and muscle movements) could provide a better understanding of the processing of multimodal emotional signals. A methodological challenge in this endeavour will be to establish perceptually-driven analyses that can describe dynamic information in both auditory and visual domains, in time-sensitive ways.

References

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current opinion in neurobiology*, 19(4), 452-458.

Audibert N., Aubergé V., Rilliard A. (2008). "How we are not equally competent for discriminating acted from spontaneous expressive speech," in *Speech Prosody 2008*, Campinan.

Aubergé, V., & Cathiard, M. (2003). Can we hear the prosody of smile?. *Speech Communication*, 40(1), 87-97.

Ay, N., Flack, J. C., & Krakauer, D. C. (2007). Robustness and complexity co-constructed in multimodal signalling networks. *Philosophical Transactions of the Royal Society B*, 362, 441-447.

Boersma, Paul & Weenink, David (2016). Praat: doing phonetics by computer [Computer program].

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

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8(4), 383-403.

Burns, K. L., & Beier, E. G. (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication*, 23(1), 118-130.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5), 1161.

RUNNING TITLE: AUTHENTICITY IN DYNAMIC AUDIOVISUAL LAUGHTER

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in cognitive sciences*, *11*(12), 535-543.

Drolet, M., Schubotz, R. I., & Fischer, J. (2013). Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 318-329.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception. *Behavioral sciences & the law*, *24*(5), 673-686.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, *38*(2), 270.

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in psychology*, *2*.

Krumhuber, E. G., & Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, *9*(6), 807.

Lampert, M. D., & Ervin-Tripp, S. M. (2006). Risky laughter: Teasing and self-directed joking among male and female friends. *Journal of Pragmatics*, *38*(1), 51-72.

Lavan, N., Lima, C. F., Harvey, H., Scott, S. K., & McGettigan, C. (2014). I thought that I heard you laughing: Contextual facial expressions modulate the perception of authentic laughter and crying. *Cognition and Emotion*, (ahead-of-print), 1-10.

Lavan, N., Scott, S. K., & McGettigan, C. (2015). Laugh Like You Mean It: Authenticity Modulates Acoustic, Physiological and Perceptual Properties of Laughter. *Journal of Nonverbal Behavior*, 1-17.

McGettigan, C., et al. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*.

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.

McKeown, G. (2015). Relevance, Mind-reading and Underdetermined Social Signals: the importance of multimodality in the psychology of human communication. Presented at the International Workshop on Advancements in Social Signal Processing for Multimodal Interaction, ACM International Conference on Multimodal Interaction ICMI, Seattle, USA.

Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology*, 31(3), 248.

Mehu, M., & van der Maaten, L. (2014). Multimodal integration of dynamic audio-visual cues in the communication of agreement and disagreement. *Journal of Nonverbal Behavior*, 38(4), 569-597.

Niewiadomski, R., Ding, Y., Mancini, M., Pelachaud, C., Volpe, G., & Camurri, A. (2015, September). Perception of intensity incongruence in synthesized multimodal expressions of laughter. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 684-690). IEEE.

Niewiadomski, R., Mancini, M., Ding, Y., Pelachaud, C., & Volpe, G. (2014, November). Rhythmic body movements of laughter. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 299-306). ACM.

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multimodal emotional stimuli?. *Motivation and Emotion*, 35(2), 192-201.

Provine, R. R. (1993). Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology*, 95(4), 291-298.

Provine, R. R. (2000). *Laughter: A scientific investigation*. Penguin.

Provine, R. R., Krosnowski, K. A., & Brocato, N. W. (2009). Tearing: Breakthrough in human emotional signaling. *Evolutionary Psychology*, 7(1), 52-56.

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

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

Wilting, J., Kraemer, E., & Swerts, M. (2006, September). Real vs. acted emotional speech. In *INTERSPEECH*.

Young, A. W., Perrett, D., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). *Facial Expressions of Emotion: Stimuli and Tests (FEEST)* Thames Valley Test Company, Bury St. Edmunds, UK.

Figure 1: Illustration of stimuli used.



Figure 2; Average ratings of authenticity for A, V, and AV stimuli of spontaneous and volitional laughter. Significant post-hoc comparisons are highlighted. Error bars show 95% confidence intervals.

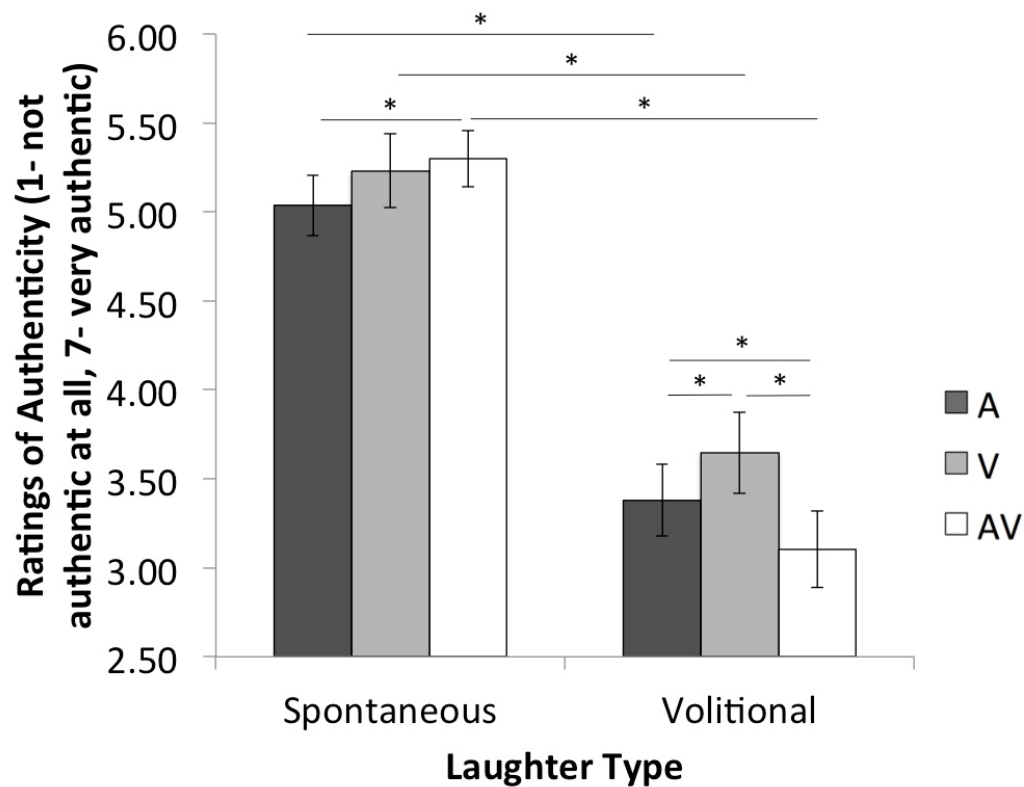


Table 1: Results of the paired sample t-tests of average ratings per participants with effect sizes, showing significant differences in the perceptual qualities of volitional and spontaneous laughter.

Modality	Scale	t	p	Cohen's d
AV	Authenticity	12.62	< .001	2.67
	Arousal	8.49	< .001	2.04
	Valence	9.48	< .001	2.68
A	Authenticity	9.14	< .001	1.73
	Arousal	11.28	< .001	1.62
	Valence	8.96	< .001	1.80
V	Authenticity	10.10	< .001	2.11
	Arousal	14.81	< .001	2.10
	Valence	8.83	< .001	2.44

***p ≤ .001, **p ≤ .01, *p ≤ .05

Table 2: Pearson's correlation coefficients per item describing the relationships between percepts of arousal, valence and authenticity.

Laughter Type	Modality	Authenticity			Valence		Arousal	
		A	AV	V	A	V	A	V
Spontaneous	AV	-	-	.620***	.830***	.299	.840***	.489*
	A	-	.863***	.443***	.908***	.110	.953***	.251
	V	-	-	-	.418*	.705***	.417*	.762***
Volitional	AV	-	-	.584***	.788***	.184	.855***	.533**
	A	-	.889**	.289	.914***	.004	.955***	.258
	V	-	-	-	.303	.698***	.296	.938***

***p ≤ .001, **p ≤ .01, *p ≤ .05

Supplementary Materials:

Means and standard deviation of ratings and difference scores of Laughterv and Laughters

Pilot Study	Modality	Scale	Laughterv	Laughters	Difference Score
-------------	----------	-------	-----------	-----------	------------------

RUNNING TITLE: AUTHENTICITY IN DYNAMIC AUDIOVISUAL LAUGHTER

			Mean	SD	Mean	SD	Mean	SD
	AV	Authenticity	2.88	0.76	4.87	0.79	1.99	0.61
		Arousal	3.53	0.78	4.98	0.70	1.46	0.66
		Valence	3.52	0.61	4.96	0.49	1.44	0.59
	A	Authenticity	3.38	0.80	4.62	0.68	1.24	0.53
		Arousal	3.72	0.70	4.76	0.64	1.04	0.36
		Valence	3.77	0.62	4.73	0.48	0.96	0.42
	V	Authenticity	3.53	0.77	5.04	0.71	1.51	0.58
		Arousal	3.74	0.64	5.13	0.73	1.39	0.36
		Valence	3.86	0.54	5.04	0.46	1.18	0.52
Main Experiment	AV	Authenticity	3.10	0.70	5.30	0.52	2.19	0.66
		Arousal	3.41	0.85	5.35	0.63	1.94	0.61
		Valence	3.08	0.76	4.94	0.70	1.87	0.65
	A	Authenticity	3.38	0.66	5.04	0.56	1.66	0.64
		Arousal	3.69	0.71	4.97	0.69	1.28	0.51
		Valence	3.37	0.67	4.74	0.72	1.37	0.56
	V	Authenticity	3.65	0.74	5.23	0.68	1.59	0.61
		Arousal	3.81	0.86	5.20	0.73	1.39	0.48
		Valence	3.44	0.75	4.88	0.80	1.44	0.62