Taylor & Francis
Taylor & Francis Group

Bioacoustics
The International Journal of Animal Sound and its Recording

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbio20

Ingressive phonation conveys arousal in human nonverbal vocalizations

Andrey Anikin & David Reby

To cite this article: Andrey Anikin & David Reby (2022) Ingressive phonation conveys arousal in human nonverbal vocalizations, Bioacoustics, 31:6, 680-695, DOI: 10.1080/09524622.2022.2039295

To link to this article: https://doi.org/10.1080/09524622.2022.2039295

Published online: 23 Feb 2022.

BIOACOUSTICS
2022, VOL. 31, NO. 6, 680–695
https://doi.org/10.1080/09524622.2022.2039295

Taylor & Francis
Taylor & Francis Group

Ingressive phonation conveys arousal in human nonverbal vocalizations

Andrey Anikin [a,b] and David Reby [b]

[a]Division of Cognitive Science, Lund University, Lund, Sweden; [b]Enes Sensory Neuro-Ethology Lab, Crnl, Jean Monnet University of Saint Étienne, St-Étienne, France

ABSTRACT

Animals normally vocalise while exhaling. Ingressive, or inspiratory, voice production occurs in humans and many other species, but its communicative function, if any, remains unknown. To test the perceptual effects of ingressive phonation, naturally occurring ingressive syllables in 109 human nonverbal vocalisations (55 laughs, 21 cries, and 33 moans) were experimentally attenuated or morphed into quiet and unvoiced intakes of breath using voice resynthesis technology. Ratings of the intensity of discrete emotions (amusement, sadness, pleasure) and of general arousal in three perceptual experiments revealed that listeners (N = 283) judged vocalisations with attenuated ingressive syllables to be less emotionally intense compared to the originals. Ingressive vocalisations were not experienced as either unnatural or unpleasant, confirming that they are a familiar part of human vocal repertoire. In sum, ingressive phonation can occur in a wide range of human nonverbal vocalisations and typically conveys intense emotion, presumably because listeners associate heavy breathing, imperfect vocal control, and continuous egressive-ingressive vocalising with the physiological state of high arousal. It remains to be seen whether ingressive phonation is a mere byproduct of high arousal or whether it can be exaggerated, and whether its communicative function extends to vocalisations of non-human animals.

ARTICLE HISTORY
Received 29 October 2021
Accepted 28 January 2022

KEYWORDS
Vocal communication; nonverbal vocalisation; emotion; arousal; ingressive; inspiratory

Introduction

The standard textbook explanation of voice production is that the lungs provide the necessary pressure, which drives a continuous airflow through the glottis and sets the vocal folds in motion (Titze 1994; Behrman 2018). As anyone who has ever gasped in surprise knows, this flow of air does not have to be in a particular direction in order to set the vocal folds vibrating, but the voice sounds rather different depending on whether we vocalise while exhaling (egressive or expiratory phonation) or inhaling (ingressive or inspiratory phonation). Ingressive phonation is generally quieter and harsher (Eklund 2008), with less energy in upper harmonics (Orlikoff et al. 1997; Vanhecke et al. 2016) and a breathy voice quality. Subjectively, it may feel effortful or unpleasant to sustain ingressive phonation for a long time

CONTACT Andrey Anikin andrey.anikin@lucs.lu.se

(Orlikoff et al. 1997) due to an anatomical asymmetry of the vocal folds (Ohala 1983), which suggests that the vocal apparatus evolved to be optimised for expiratory airflow. This design feature makes sense physiologically: it is easier to achieve high pressure when contracting, as opposed to enlarging, the rib cage (Behrman 2018). Furthermore, delaying an intake of breath with a long vocalisation may be more costly metabolically compared to prolonging the exhalation. As a result, humans, and apparently most other mammals, find it more natural to vocalise during exhalation.

While ingressive phonation is clearly the exception rather than the rule, the list of these exceptions has been steadily growing without any unifying theoretical framework to explain them. A wide range of animal vocalisations are partly ingressive, including pant-hoots of chimpanzees (Riede et al. 2004), laughs of all great apes (Ross et al. 2009), purring of felines (Eklund et al. 2010), songs of gibbons (Geissmann 1984) and birds (Goller and Daley 2001), bellows of koalas (Charlton 2015), groans of fallow deer (Reby and McComb 2003), display calls of African penguins (Favaro et al. 2014), braying of donkeys, roars of howler monkeys, etc. In humans, ingressive phonation is found in some speech registers (Eklund 2008) and in singing (Vanhecke et al. 2016), but above all in nonverbal vocalisations such as laughs (Tanaka and Campbell 2011; Bryant 2020), sobs (Darwin 1872; Aucouturier et al. 2011), and gasps (Anikin 2020a). We are not aware of any quantitative analyses of the prevalence of ingressive phonation in vocalisations of adult humans, but it was straightforward to find over a hundred examples with loud ingressive syllables for this study, suggesting that they are fairly common. This raises the following question: is the presence of ingressive phonation in nonverbal vocalisations an epiphenomenon, or does it contribute meaningfully to communication? There are two theoretical considerations that suggest a possible function for ingressive phonation.

First, polysyllabic bouts of vocalising unfold over multiple respiratory cycles and therefore require precisely timing voice onsets and offsets to keep them purely egressive, particularly when the breathing is heavy and rapid due to generally high arousal. In other words, because respiratory rate increases and vocal control becomes more challenging with mounting physiological arousal, ingressive syllables may be increasingly difficult to avoid, making them reliable indicators of the vocaliser's agitated state. To take an extreme example, respiration may be seriously disrupted during a fit of uncontrollable giggles, leaving the sufferer literally – and audibly – gasping for breath. As a result, listeners can be expected to interpret heavily ingressive laughs as a sign of genuine, intense amusement. This reasoning primarily applies to rapid egressive-ingressive call sequences, such as sobbing in humans or pant-hoots in chimpanzees (Riede et al. 2004), but probably not to call bouts that are produced at a relatively slow rate (e.g. purring of cats, human moans) or to purely ingressive calls (e.g. isolated human gasps or fallow deer groans).

Second, phonating during both exhalation and inhalation enables an animal to maximise the number of calls per unit of time, as well as the proportion of time spent vocalising, which are important acoustic indicators of high arousal (Volodin

et al. 2009; Briefer 2012). Unlike players of wind instruments who have mastered circular breathing (White 2014), most ordinary mammals have to interrupt their vocal production now and then to draw a breath. Ingressive phonation can fill the silences otherwise left between syllables, increasing the rate of acoustic events and intensifying the 'acoustic bombardment' of listeners. This is important because intense vocalisations, and in fact emotional speech, need to be effective at involuntarily attracting and holding the attention of listeners, which creates a close alignment between emotion intensity and low-level auditory salience of vocalisations (Anikin 2020b). In other words, it may be adaptive for callers to vocalise during inhalation in urgent, high-arousal contexts, such as intense distress, in order to maximise the salience of their vocal output.

Both theoretical arguments lead to the same prediction: ingressive phonation should perceptually convey high arousal or emotion intensity. Two approaches can be used to test this prediction. One is to correlate the degree of ingressiveness in naturally occurring vocalisations with other perceptual qualities. In the only such study that we are aware of, Kret et al. (2021) found that subjective ratings of ingressiveness were negatively correlated with the pleasantness or contagiousness of infant laughter, concluding that infants are socialised to laugh in a purely egressive manner because adults find this 'proper' laughter more pleasing. While the above study is a welcome pioneering attempt to shed some light on the role of ingressive phonation in laughter, the analysed laughs presumably differed in many acoustic characteristics apart from their perceived ingressiveness. A stronger case can be made if vocalisations are manipulated experimentally to change the amount of ingressive phonation while preserving their temporal structure, pitch, voice quality, and other acoustic properties. This is the approach taken in the current study: nonverbal vocalisations with a loud, voiced ingressive syllable had this syllable attenuated and morphed into inhalation noise, turning it into a soft intake of breath with approximately the same spectral and amplitude envelopes as in the original. The stimuli consisted of three types of polysyllabic human nonverbal vocalisations that often contain ingressive syllables: laughs, cries or sobs, and moans.

Laughter is extremely common in everyday life, present in all human societies (Provine 2000; Bryant et al. 2018), found in children born deaf and blind (Eibl-Eibesfeldt 1989), and similar to the laughter of great apes in terms of its function and basic acoustics (Ross et al. 2009). The evolutionary origins of laughter are assumed to lie in rough-and-tumble social play accompanied by laborious breathing, which was eventually ritualised into a pant-like vocalisation (van Hooff 1972; Provine 2000; Vettin and Todt 2005; Bryant 2020). Consistent with this hypothesis, great apes produce pant-like laughs with alternating egressive and ingressive syllables. Interestingly, human infants initially laugh in an 'ape-like', egressive-ingressive manner and only switch to adult-type, primarily egressive laughter over the first few years of life (Kret et al. 2021). Notwithstanding this developmental trajectory, the acoustic variability of laughs should not be underestimated (Bachorowski et al. 2001; Vettin and Todt 2004), and ingressive phonation is well documented in the laughter of adult humans (Eklund 2008; Tanaka and Campbell 2011; Bryant 2020).

Crying in humans is functionally similar to distress calls in other animals and has some acoustic parallels with the screams and whimpers of any young mammal (Lingle et al. 2012). However, human crying is a multimodal nonverbal behaviour with various components emerging at different times in development. Newborn infants are already capable of producing screams and cries of distress of the kind analysed by Lingle et al. (2012), but both weeping with tears and egressive-ingressive sobbing emerge only after a few months (Darwin 1872). Crying in adults is less common than in infants, but very distinct acoustically when it does occur (Anikin et al. 2018). Ingressive phonation in crying appears to be very common, as captured in the English word to sob.

The third vocalisation type we tested is moaning. There has been little formal research on moaning, but moans are known to be common in painful contexts such as childbirth, as well as during sexual activities (Anikin and Persson 2017; Prokop 2021), and are sometimes joined by loud ingressive syllables. Gasps were not considered because these single-syllable calls are fully ingressive by definition (Anikin 2020a).

For all three call types (laughs, cries, and moans), experimentally attenuating their ingressiveness was predicted to lower the perceived level of arousal and emotion intensity. To test this hypothesis, listeners rated the original and manipulated vocalisations on several perceptual scales designed to capture the perceived level of general arousal, the intensity of expressed emotion (amusement for laughs, sadness for cries, and pleasure or pain for moans), and induced emotion (contagiousness, pleasing or disturbing nature of manipulated vocalisations). Finally, authenticity ratings were obtained to test whether removing ingressive phonation made the vocalisations more or less natural.

Methods

Stimuli

Recordings of laughs, cries, and moans (Table 1) were chosen for high audio quality and the prominence of ingressive phonation. Laughs (n = 55) were obtained from two sources: 21 from a collection of nonverbal vocalisations from YouTube videos (Anikin and Persson 2017) and 34 from unscripted dyadic interactions based around retelling funny videos (Wood 2020). In both collections, adult humans were laughing spontaneously and presumably were genuinely amused, although the acoustic intensity of laughter was considerably higher in the YouTube collection. Each clip was trimmed to contain a single prominent and at least partly voiced ingressive syllable, which could be located at the onset (n = 8), offset (n = 7), or in the middle (n = 40) of the laugh. Ingressive syllables were manually annotated; their duration was (mean ± SD) 436 ± 152 ms, range [237, 1086].

Cries of sadness (n = 21) were taken from Anikin and Persson (2017) and trimmed to contain 1–3 ingressive syllables. Six were produced by school-age children, and 15 by adult men and women. Moans (n = 33) were taken from recordings of women giving birth (n = 16 recordings) and engaging in real or simulated sexual activities (n = 17), which were obtained from two sources (Magnard 2014; Anikin and Persson 2017) and edited to contain between one and three voiced ingressive syllables and one or more egressive syllables.

Table 1. Descriptives of stimuli and perceptual rating scales.

	Laughs	Cries	Moans
N	55 (30 female, 25 male)	21 (11 female, 4 male, 6 child)	33 (female)
Duration, s mean ± SD [range]	2.2 ± 0.6 [0.9, 3.3]	2.6 ± 1.2 [1.5, 6.8]	3.1 ± 1.0 [1.5, 5.0]
Perceptual rating scales	*Amusement* How amused does this person sound? Is something amusing, funny going on?	*Sadness* How sad does this person sound? Is something distressing, sad going on?	*Valence* Does this sound like pain or pleasure? Some moans were recorded in childbirth, others in sexual contexts.
	Emotion intensity How agitated/excited does this person sound? Please indicate whether the person sounds calm or agitated/excited.		
	Contagion How contagious do you find this laugh? Do you feel like joining in when you hear this laugh?	*Contagion* How contagious do you find this vocalisation? Do you get sad/feel like crying yourself when you hear this sound?	*Contagion* How contagious do you find this vocalisation? Do you feel emotional yourself when you hear this sound?
	Pleasantness How pleasant do you find this laugh? Please indicate whether you find this laugh pleasant or unpleasant to hear.	*Disturbing quality* How disturbing do you find this vocalisation? Please indicate whether you find this cry disturbing, unpleasant to hear.	*Disturbing quality* How disturbing do you find this vocalisation? Please indicate whether you find this moan disturbing, unpleasant to hear.
	Authenticity How authentic/realistic do you find this vocalisation? Does this sound realistic or natural, as something you might hear in real life?		

Manipulation of ingressiveness

Ingressive syllables were manually annotated and manipulated in one of two ways. In the *original* condition, they were attenuated by 0, 9, or 18 dB by means of separating, processing, and then cross-fading egressive and ingressive parts of each laugh with an R script. In the noise condition, there was an additional processing step: the ingressive syllable was mixed with a synthetic noise with the same RMS amplitude and the same spectral envelope, which was achieved by using the *transplantFormants()* function in the R package *soundgen* (Anikin 2019). The relative contribution of noise was 0% at attenuation = 0 dB (i.e. unmodified original recording), 50% at attenuation = 9 dB, and 100% at attenuation = 18 dB (i.e. only noise). In other words, in addition to making the ingressive syllable quieter, in the *noise* condition it was morphed into an aspiration-like sound with the same formant structure. Obviously, this is only an approximation since an unvoiced inhalation would not have exactly the same smoothed spectral envelope as a voiced ingressive syllable. Coupled with the manipulation of intensity, however, in most cases this produced a passable impression of the speaker drawing a quiet breath without phonating, while the egressive parts of the original vocalisation remained unchanged (Figure 1).

Participants

Participants (N = 283) were recruited on the online testing platform Prolific (https://www.prolific.co/). Independent samples of listeners rated laughs (n = 100 listeners), cries (n = 82 listeners), and moans (n = 101 listeners). All participants

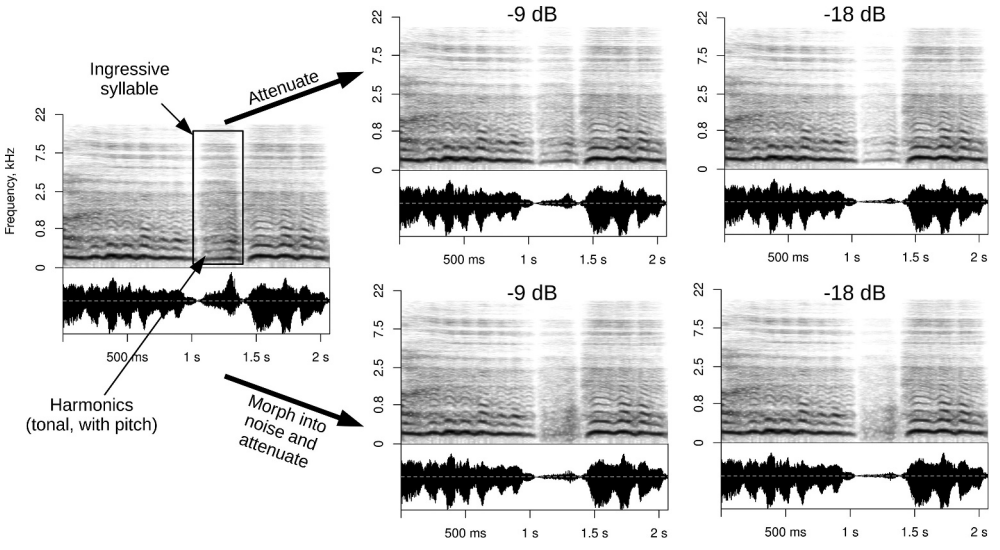

Figure 1. Experimental manipulations of the ingressive syllable, here shown for one of 55 prototype laughs. Spectrograms use the bark frequency scale for effective visualisation of both lower harmonics and formants.

self-reported to have normal hearing and to be fluent in English; 55% self-identified as female, 43% as male, and 1% as 'unspecified'; the average age was 24 ± 6 years, range [18, 59]. Sample sizes were chosen to ensure sufficient precision of estimates of population-level effect sizes in Bayesian multilevel models. This precision primarily depends on the number of stimulus prototypes (here 55 + 21 + 33 = 109) and the number of times each unique manipulated sound (here 109 * 6 = 654) is tested on each response scale (here 6.7 times for laughs and moans, 8.3 times for cries). The average width of 95% credible intervals (CIs) of effect sizes was 4.0% for laughs, 6.5% for cries, and 6.9% for moans, and the main factor limiting the precision of this analysis is the number of prototypes available, rather than the number of raters.

Procedure

We obtained listeners' ratings of the manipulated stimuli on five response scales (Table 1) in three online experiments (for laughs, cries, and moans), each with an independent sample of raters. Each participant rated all prototypes in two (laughs and moans) or three (cries) blocks, each with a randomly selected response scale. The order of blocks and trials within blocks was randomised for each participant. Responses were given on a horizontal Visual Analog Scale labelled at the extremes, without tick marks but with grey stripes.

Data analysis

Each of the three experiments (laughs, cries, and moans) was analysed independently. Unaggregated responses were analysed with Bayesian multilevel zero-one-inflated beta models using the R package *brms* (Bürkner 2017). The model predicted

the rating as a function of Condition (2 levels: original or noise), Attenuation of ingressive syllable (linear effect: 0, 9, or 18 dB), and Scale (5 levels), with all possible interactions. Because an attenuation of 0 dB means that the sounds were identical in both conditions, the corresponding beta coefficients were set to zero as part of prior specification, forcing the separate regression lines per condition to converge at zero (Figure 2). The effects of condition and scale were assumed to vary across subjects and across prototypes, and the effect of scale could also vary across individual

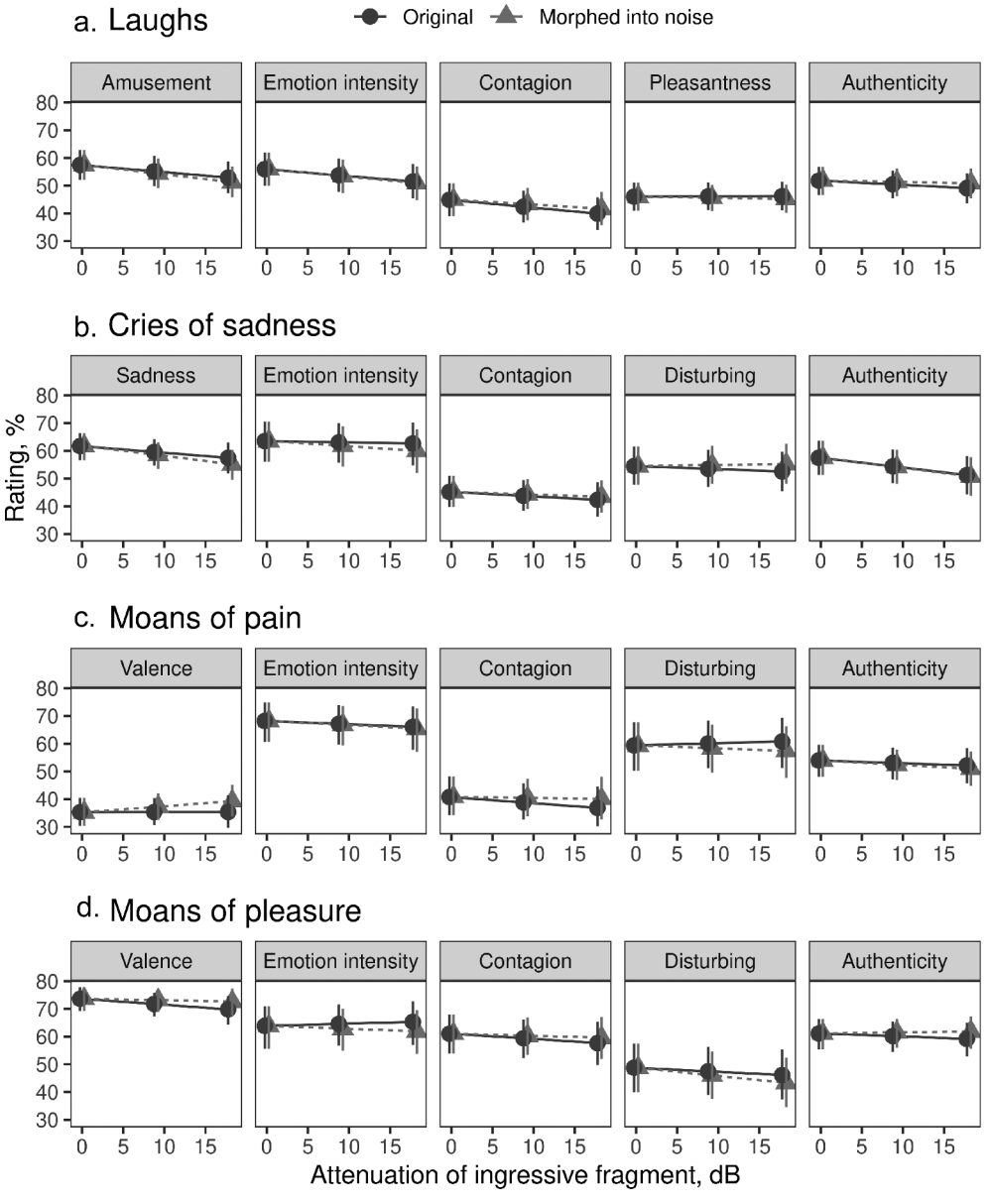

Figure 2. Perceptual ratings of laughs (A), cries (B), moans of pain (C), and moans of pleasure (D): fitted values from mixed models and 95% CI.

stimuli. Finally, the variance of responses (phi) was assumed to vary across participants to account for individual differences in using the response scales. The model structure in *brms* syntax was as follows:

response ~ condition * attenuation * scale + (condition * scale | subject + prototype) + (scale | stimulus), phi ~ (scale | subject)

As a measure of inter-rater agreement in the rating task, we aggregated the ratings of each vocal stimulus on each response scale and calculated the mean Pearson's correlation between the responses of each participant and these aggregated ratings. Averaging across five response scales, this correlation was r = .65 for laughs, .53 for cries, and .63 for moans. Comparing different scales, the lowest inter-rater reliability was obtained when rating the contagiousness of cries (r = .38) and the authenticity of moans (r = .39), suggesting that these judgements were rather idiosyncratic.

Posterior distributions of model parameters and fitted values were summarised by their medians and 95% credible intervals (CIs). The proportion of posterior distribution (PD) of effect sizes that is positive (or negative) is also reported in the text. The audio, datasets, and R code for audio manipulation and data analysis are available in online supplements at http://cogsci.se/publications.html.

Results

The effect sizes below correspond to fitted values of the difference between manipulated and unmanipulated stimuli on a scale of 0 to 100, reported as percentage points. The effects mentioned in the text are for the maximum attenuation of 18 dB (see Figures 2 and 3 for full results), which turned a loud ingressive syllable into a barely audible one in the original condition, or into a quiet intake of breath in the noise condition.

Laughs

Amusement, arousal, and contagion scales were correlated (all r > .71), and all three showed a robust negative effect of attenuating ingressive syllables (Figure 3A). Amusement ratings became 4.5% lower (95% CI [1.9, 7.0], 100% of posterior distribution [PD] > 0) in the original condition and 6.2% [3.5, 8.8] lower (100% of PD > 0) in the noise condition. Similar effect sizes were observed for emotion intensity (4.5% [2.1, 6.7] and 5.1% [2.8, 7.3]; 100% of PD > 0 for both) and contagion (3.2% [0.7, 5.9] and 5.0% [2.6, 7.4]; 99.2 and 100% of PD > 0, respectively). Thus, ingressive laughs were experienced as more intense and contagious.

Attenuating ingressive syllables did not noticeably change the pleasantness of laughs: by 0.2% [−2.6, 3.2] in the *original* condition and −0.8% [−3.8, 2.2] in the *noise* condition. There was a slight, statistically uncertain tendency for laughs to sound less authentic when the original ingressive syllable was made quieter (by 2.6%, 95% CI [0.5, 5.8], 95.0% of PD > 0), but no change at all in perceived authenticity when it was morphed into breathing noise (−0.9% [−3.8, 2.0], 72.2% of PD < 0). Thus, experimental manipulations did not sound unnatural, but neither did laughs become more authentic when loud ingressive syllables were replaced with quiet breathing.

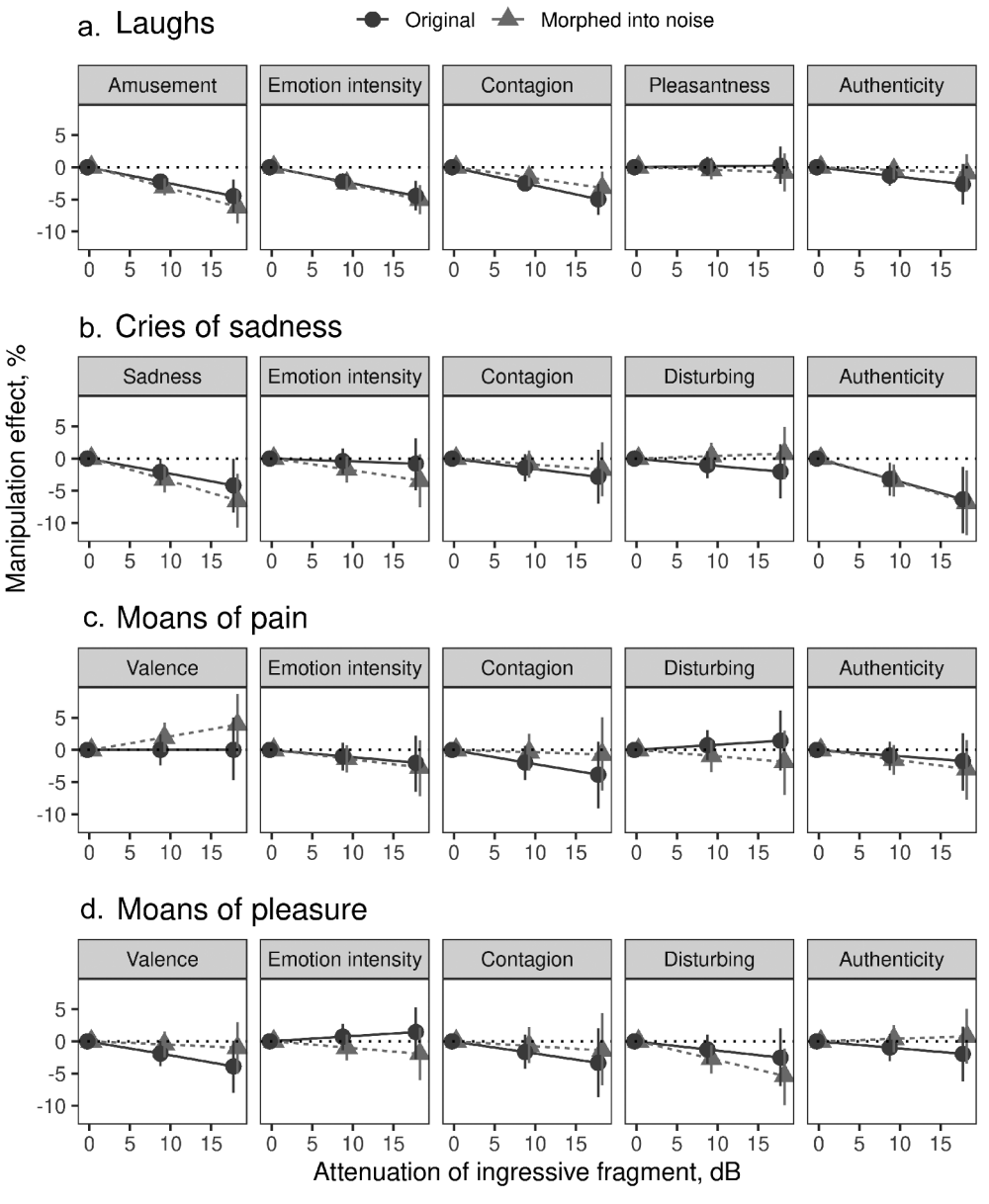

Figure 3. The effect of experimentally attenuating ingressive syllables on perceptual ratings of laughs (A), cries (B), moans of pain (C), and moans of pleasure (D). Solid points are medians of posterior distributions of the difference in fitted values between each condition and the original (attenuation = 0 dB), with 95% CI.

Cries

Cries whose ingressive syllables were turned into quiet breathing were rated lower on sadness (6.5% [2.4, 10.7], 99.9% of PD > 0) and authenticity (6.8% [1.8, 11.9], 99.6% of PD > 0). There was also a statistically less certain tendency for them to be rated lower on emotion intensity (3.4% [−0.7, 7.5], 95.2% of PD > 0). The effects of simply making the

ingressive syllables quieter were qualitatively similar, but less pronounced compared to morphing them into breathing (Figure 3). Thus, less ingressive cries were perceived as less intense emotionally, but also less natural or authentic-sounding, even when ingressive syllables were simply made quieter.

Moans

Whereas laughing is a relatively unambiguous signal of amusement or joy, and crying of sadness, moans can express very different emotions (Anikin et al. 2018). Among the 33 tested moans, 17 were moans of sensual pleasure, and all 17 had mean valence ratings above 50% before manipulations of their ingressiveness (mean = 77%), suggesting that they were recognised as pleasure (Figure 2C-D). Likewise, 15 out of 16 childbirth moans had a mean valence rating under 50% (mean = 32%), so the context was usually recognised as unpleasant, but with more variability compared to sexual moans. There may be differences in how moans with positive and negative perceived valence are affected by the manipulations of ingressiveness. For example, if less ingressive moans sound less emotionally intense, this would be expected to translate into lower valence ratings for moans of pleasure (less intense pleasure = less positive valence), but higher valence for moans of pain (less intense pain = more positive valence). Accordingly, moans were analysed with an additional interaction with their production context: pain (giving birth) or pleasure (sexual activities).

No statistically robust effects of manipulating ingressiveness were observed for moans of pain (Figure 3C), except for an uncertain tendency for less negative valence in the *noise* condition (3.9% [−0.7, 8.7], 95.1% of PD > 0) after ingressive syllables were morphed into quiet breathing (Figure 3D). In contrast, moans of pleasure became less disturbing in the *noise* condition (5.4% [1.1, 9.9], 99.5% of PD > 0) and less positive in valence in the *original* condition (3.9% [−0.1, 8.0], 97.2% of PD > 0). In other words, moans tended to become more neutral in valence when ingressive syllables were attenuated.

Discussion

We used voice resynthesis technology to directly manipulate the ingressiveness of human laughs, cries, and moans, revealing that ingressive phonation enhances the perceived level of emotion intensity. Thus, morphing loud ingressive syllables into quiet, unvoiced breaths makes it appear that a person laughing is less amused, a person crying is less sad, and a person moaning is not enjoying herself as much as when audible ingressive phonation is preserved in otherwise identical nonverbal vocalisations. This confirms the hypothesis that the presence of ingressive phonation conveys high arousal. Below, we discuss the findings in relation to the three tested call types – laughs, cries, and moans – and consider their implications for the study of animal and human vocal communication.

The presence of ingressive gasp-like syllables in laughs was found to enhance the perceived level of amusement, emotion intensity, and contagiousness of laughs. This agrees very well with the notion of the 'animal nature' of spontaneous laughter advocated by Greg Bryant and Aktipis (2014), Bryant (2020)), who argues that regular, completely voiced laughs are produced under the control of the neurological circuits responsible for speech and are therefore perceived as less spontaneous, whereas laughs caused by

a powerful, genuine emotion rely on evolutionarily older brain stem mechanisms and are more variable acoustically. Specifically, rapid heavy breathing associated with high physiological arousal is likely to introduce noisy and occasionally ingressive syllables into laughs (Gervais and Wilson 2005). Accordingly, listeners are probably correct in interpreting ingressive phonation in laughs as a signal of high arousal and intense amusement.

Contrary to the recently published evidence that ingressiveness ratings of infant laughs negatively correlated with their pleasantness and contagiousness ratings (Kret et al. 2021), ingressive laughs in the present study were not experienced as either unnatural or unpleasant, and in fact were rated as more, not less contagious. Direct manipulation of the target acoustic characteristic is a more powerful method than the correlation analysis used by Kret et al. (2021). On the other hand, the manipulated laughs contained only a single ingressive syllable; repeated ingressive syllables in a long bout of laughing might indeed begin to sound irritating. Intense nonverbal vocalisations are seldom pleasant to listen to – in fact, intense delight may sound remarkably similar to extreme anguish (Anikin and Persson 2017; Atias et al. 2019; Atias and Aviezer 2020). Nevertheless, it appears unlikely that the negative correlation between ingressiveness and pleasantness or contagiousness reported by Kret et al. (2021) in infant laughs is a general phenomenon as there was no increase in contagiousness or pleasantness ratings of any vocalisations when their ingressiveness was experimentally reduced. Ingressive phonation thus appears to be a normal feature of human laughter, neither uncommon nor repulsive to listeners. Accordingly, it is probably an exaggeration to draw a strict distinction between egressive-ingressive laughter of other apes and egressive-only laughter of adult humans, as suggested in some publications (e.g. Provine 2000; Ross et al. 2009; Kret et al. 2021).

As expected, cries that were experimentally made less ingressive conveyed less intense sadness. Interestingly, they were also rated as less authentic, even when the ingressive syllables were simply made quieter without rendering them unvoiced. As no such effect was observed in laughs and moans, ingressiveness may be such a common feature in intense sobbing that its absence is experienced as unnatural. If so, ingressive phonation may need to be added to the list of acoustic features that 'make a cry a cry' (Lingle et al. 2012) alongside chevron-shaped pitch contours and nonlinear phenomena. Only cries of adults and school-age children were tested in this study. Considering the higher prevalence of crying in infants, as well as the great theoretical and practical importance of infant vocalisations, in future studies it will be important to extend the present findings by examining ingressive phonation in baby cries. It is well known that baby cries contain ingressive syllables (Darwin 1872; Aucouturier et al. 2011), but their prevalence and communicative significance have not been investigated. Judging by the present findings, ingressive phonation may turn out to be an important marker of intense distress in the crying of infants.

Moans are different from laughs and cries in their lack of an intrinsic temporal structure: a single vocalisation is normally produced per respiratory cycle, with no apparent sequences comparable to bouts of laughing or crying. It is therefore possible that listeners are less familiar with ingressive syllables between moans than they are with ingressive laughs and cries. Manipulation effects were less clear-cut for moans compared to laughs and cries, but the observed patterns are consistent with the hypothesis that

ingressiveness in moans may intensify the conveyed level of pain or pleasure. In particular, moans became more neutral in valence when we attenuated voiced ingressive syllables between them. Interestingly, breathiness was previously found to be associated with perceived intensity of pleasure in moans (Anikin 2020a), so there appears to be a general expectation to find breathy and ingressive moans in erotic contexts, which can be investigated further in future studies.

The distinction between spontaneous and volitional vocal production (Bryant and Aktipis 2014; Anikin and Lima 2018; Atias and Aviezer 2020) is important to consider when studying acoustic indicators of emotion intensity such as ingressiveness. Humans can exert volitional control over their voice and imitate or modify otherwise innate vocalisations such as laughs (Ackermann et al. 2014), which means that speakers may be suppressing or exaggerating ingressiveness and other acoustic features in accordance with the context and cultural expectations. For instance, while the ingressiveness of 'animal-like' spontaneous laughter may be involuntary and indexical of high arousal, ingressiveness in erotic moans might well be exaggerated volitionally to conform to cultural stereotypes or please the partner. In fact, while 68% of women in a recent survey (Prokop 2021) reported moaning during intercourse, 38% also reported faking sexual vocalisations. Like other indicators of emotional arousal, ingressiveness is thus open to intentional manipulation.

The novel method of manipulating ingressiveness that was developed for this study produced encouraging results. Simply making ingressive syllables quieter (by separating, processing, and then cross-fading separate syllables) is a viable 'low-tech' approach, but the alternative method of morphing voiced ingressive syllables into aspiration noise consistently produced stronger effects, while preserving the naturalness of vocalisations. It works particularly well for removing the tonal component from relatively short ingressive syllables surrounded by egressive syllables – for example, in egressive-ingressive laughs. This manipulation sounds highly authentic because both the spectral envelope (formant structure) and the amplitude envelope of the original are preserved, but some adjustment of overall amplitude is still necessary because breathing is seldom as loud as ingressive phonation. It is more technically challenging to turn very high-pitched or whistle-like ingressive sounds into aspiration, and especially to add ingressive syllables where they were none originally, which is why in this study we only worked with original vocalisations that already contained ingressive syllables. Despite these current limitations, the ability to manipulate ingressiveness directly is a powerful experimental technique and another example of the benefits of using voice synthesis technology in research on vocal communication.

An important limitation of the present study is that only perceptual effects of manipulating ingressiveness were tested. In future it will be essential to compare the ingressiveness of laughs and other vocalisations produced under low and high arousal, verifying that ingressive phonation indeed occurs more frequently as the urgency or intensity of emotional state increases. A good start would be to perform a large-scale, quantitative analysis of the prevalence of ingressive phonation in human and animal nonverbal vocalisations associated with different contexts, which can actually be challenging to do because it is not always clear which syllables are ingressive. Targeted manipulation of ingressiveness, such as the method proposed in the present study, in combination with perceptual experiments in humans and playback studies in other species can then provide decisive evidence on the role that ingressive phonation plays in vocal communication.

To summarise, the present results demonstrate experimentally that vocalising during both exhaling and inhaling – egressively and ingressively – conveys high arousal in a range of human nonverbal vocalisations such as laughing, crying, and moaning. Assuming that this perceptual bias has some foundation in reality – that is, that ingressive phonation is indeed more common when vocalising in a state of high physiological arousal – an important area for future research will be to investigate the extent to which ingressiveness is a mere consequence of vocalising in an excited state and/or a flexible acoustic feature that may be exploited and exaggerated by speakers. On the one hand, ingressive syllables may be considered 'vocal slips' that betray a lack of vocal control: as the breathing rate increases and repeated vocalisations are produced under high arousal, imperfect timing of voice onsets and offsets may create unintentional gasps, whistles, wheezes, gurgles, coughs, and other unexpected sounds that introduce such incredible variety into high-intensity spontaneous vocalisations such as laughs (Bachorowski et al. 2001; Vettin and Todt 2004; Bryant 2020). Ingressive syllables may then function as indexical and therefore honest (Bradbury and Vehrencamp 2011) signals of intense emotion. On the other hand, it may be adaptive for the caller to utilise both exhalation and inhalation for vocalising, as this would optimise the production of acoustic events per unit of time. From this perspective, and perhaps especially in slower vocalisations such as sequences of moans, ingressive syllables may be introduced more or less voluntarily in order to maximise auditory salience and hold the listeners' attention. To learn more about the evolution and diverse roles of ingressive phonation, it will be crucial to systematically investigate its occurrence and communicative functions in vocalisations of non-human animals.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

A.A. was supported by grant Vetenskapsrådet [2020-06352] from the Swedish Research Council. D.R. was supported by the University of Lyon IDEXLYON project as part of the 'Programme Investissements d'Avenir' (ANR-16-IDEX-0005).

ORCID

Andrey Anikin http://orcid.org/0000-0002-1250-8261
David Reby http://orcid.org/0000-0001-9261-1711

Ethical statement

All participants provided informed consent. Ethical approval for performing perceptual experiments with human subjects was provided by the Comité d'Ethique du CHU de Saint-Etienne (IRBN692019/CHUSTE).

Data accessibility

All datasets, R scripts for data analysis, audio stimuli, R scripts for generating synthesized stimuli, and html code for running the psychoacoustic experiment can be downloaded from http://cogsci.se/publications.html. These supplemental materials enable full validation and replication of results.

Authors' contributions

A.A. and D.R conceived and designed the studies. A.A. created the vocal stimuli, programmed the experiments, collected and analyzed the data, and drafted the paper. A.A. and D.R. wrote the final manuscript, and all authors approved the final version.

References

Ackermann H, Hage SR, Ziegler W. 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. Behav Brain Sci. 37 (6):529–546. doi:10.1017/S0140525X13003099.

Anikin A. 2018. Soundgen: an open-source tool for synthesizing nonverbal vocalizations. Behav Res Methods. 1–15. doi:10.3758/s13428-017-1001-8.

Anikin A. 2020a. A moan of pleasure should be breathy: the effect of voice quality on the meaning of human nonverbal vocalizations. Phonetica. 77(5):327–349. doi:10.1159/000504855.

Anikin A. 2020b. The link between auditory salience and emotion intensity. Cogn Emot. 34 (6):1246–1259. doi:10.1080/02699931.2020.1736992.

Anikin A, Bååth R, Persson T. 2018. Human non-linguistic vocal repertoire: call types and their meaning. J Nonverbal Behav. 42(1):53–80. doi:10.1007/s10919-017-0267-y.

Anikin A, Lima C. 2018. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. Q J Exp Psychol. 71(3):622–641.

Anikin A, Persson T. 2017. Nonlinguistic vocalizations from online amateur videos for emotion research: a validated corpus. Behav Res Methods. 49(2):758–771. doi:10.3758/s13428-016-0736-y.

Atias D, Aviezer H. 2020. Real-life and posed vocalizations to lottery wins differ fundamentally in their perceived valence. Emotion. Advance online publication. 10.1037/emo0000931

Atias D, Todorov A, Liraz S, Eidinger A, Dror I, Maymon Y, Aviezer H. 2019. Loud and unclear: intense real-life vocalizations during affective situations are perceptually ambiguous and contextually malleable. J Exp Psychol General. 148(10):1842. doi:10.1037/xge0000535.

Aucouturier -J-J, Nonaka Y, Katahira K, Okanoya K. 2011. Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. J Acoust Soc Am. 130(5):2969–2977. doi:10.1121/1.3641377.

Bachorowski J-A, Smoski MJ, Owren MJ. 2001. The acoustic features of human laughter. J Acoust Soc Am. 110(3):1581–1597. doi:10.1121/1.1391244.

Behrman A. 2018. Speech and voice science. 3rd ed. San Diego: Plural publishing.

Bradbury JW, and Vehrencamp SL. 2011. Principles of animal communication. 2nd. Sinauer (Sunderland, MA), Sinauer Associates, Sunderland.

Briefer EF. 2012. Vocal expression of emotions in mammals: mechanisms of production and evidence. J Zool. 288(1):1–20. doi:10.1111/j.1469-7998.2012.00920.x.

Bryant GA. 2020. Evolution, structure, and functions of human laughter. In: Floyd F, Weber R, editors. The handbook of communication science and biology. New York: Routledge; p. 63–77.

Bryant GA, Aktipis CA. 2014. The animal nature of spontaneous human laughter. Evol Hum Behav. 35(4):327–335. doi:10.1016/j.evolhumbehav.2014.03.003.

Bryant GA, Fessler DM, Fusaroli R, Clint E, Amir D, Chávez B, Zhou Y. 2018. The perception of spontaneous and volitional laughter across 21 societies. Psychol Sci. 29(9):1515–1525. doi:10.1177/0956797618778235.

Bürkner P-C. 2017. brms : an R Package for Bayesian multilevel models using stan. J Stat Softw. 80 (1):1–28. doi:10.18637/jss.v080.i01.

Charlton BD. 2015. The acoustic structure and information content of female koala vocal signals. PloS One. 10(10):e0138670. doi:10.1371/journal.pone.0138670.
Darwin C. 1872. The expression of emotions in man and animals. 1899 ed. New York: Appelton.
Eibl-Eibesfeldt I. 1989. Human ethology. New York: Aldine de Gruyter.
Eklund R. 2008. Pulmonic ingressive phonation: diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. J Int Phon Assoc. 38(3):235–324. doi:10.1017/S0025100308003563.
Eklund R, Peters G, Duthie ED. 2010. An acoustic analysis of purring in the cheetah (Acinonyx jubatus) and in the domestic cat (*Felis catus*). In Fonetik 2010, Lund University. Lund (Sweden): Mediatryck, Lund, p. 17–22.
Favaro L, Ozella L, Pessani D. 2014. The vocal repertoire of the African Penguin (Spheniscus demersus): structure and function of calls. PloS One. 9(7):e103460. doi:10.1371/journal.pone.0103460.
Geissmann T. 1984. Inheritance of Song Parameters in the Gibbon Song, Analysed in 2 Hybrid Gibbons *(Hylobates pileatus* × *H. lar)*. Folia Primatol. 42(3–4):216–235. doi:10.1159/000156165.
Gervais M, Wilson DS. 2005. The evolution and functions of laughter and humor: a synthetic approach. Q Rev Biol. 80(4):395–430. doi:10.1086/498281.
Goller F, Daley MA. 2001. Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. Proc BBiol Sci. 268(1483):2301–2305. doi:10.1098/rspb.2001.1805.
Kret M, Venneker D, Evans B, Samara I, Sauter D. 2021. The ontogeny of human laughter. Biol Lett. 17(9):20210319. doi:10.1098/rsbl.2021.0319.
Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA. 2012. What makes a cry a cry? A review of infant distress vocalizations. Curr Zool. 58(5):698–726. doi:10.1093/czoolo/58.5.698.
Magnard C 2014. Caracteristiques acoustiques des cries de la femme a l'accouchement: une etude preliminaire prospective chez 16 patientes. [Acoustic characteristics of childbirth cries: a preliminary prospective study in 16 patients.] Unpublished doctoral thesis. Universite de Saint-Etienne, France. French.
Ohala J. 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage PF, editors. The production of speech. New York: Springer; p. 189–216.
Orlikoff RF, Baken RJ, Kraus DH. 1997. Acoustic and physiologic characteristics of inspiratory phonation. J Acoust Soc Am. 102(3):1838–1845. doi:10.1121/1.420090.
Prokop P. 2021. Factors influencing sexual vocalization in human females. Arch Sex Behav. 50(8):3809–3820. doi:10.1007/s10508-021-02018-4.
Provine RR. 2000. Laughter: a scientific investigation. New York: Penguin Press.
Reby D, McComb K. 2003. Vocal communication and reproduction in deer. Adv Study Behav. 33:231–264.
Riede T, Owren MJ, Arcadi AC. 2004. Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): frequency jumps, subharmonics, biphonation, and deterministic chaos. Am J Primatol. 64(3):277–291. doi:10.1002/ajp.20078.
Ross MD, Owren MJ, Zimmermann E. 2009. Reconstructing the evolution of laughter in great apes and humans. Curr Biol. 19(13):1106–1111. doi:10.1016/j.cub.2009.05.028.
Tanaka H, Campbell N. 2011. Acoustic features of four types of laughter in natural conversational speech. ICPhS. 1958–1961.
Titze I. 1994. Principles of voice production. Englewood Cliffs (NJ): Prentice Hall.
van Hooff J. 1972. A comparative approach to the phylogeny of laughter and smiling. In: Hinde RA, editor. Nonverbal communication. Cambridge University Press; p. 209–241.
Vanhecke F, Lebacq J, Moerman M, Manfredi C, Raes G-W, DeJonckere PH. 2016. Physiology and acoustics of inspiratory phonation. J Voice. 30(6):769–e9. doi:10.1016/j.jvoice.2015.11.001.
Vettin J, Todt D. 2004. Laughter in conversation: features of occurrence and acoustic structure. J Nonverbal Behav. 28(2):93–115. doi:10.1023/B:JONB.0000023654.73558.72.
Vettin J, and Todt D. 2005. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. Behav. 142(2):217–240.

Volodin IA, Volodina EV, Gogoleva SS, Doronina LO. 2009. Indikatory emotsionalnogo vozbuzhdeniya v zvukah mlekopitayuschih i cheloveka. [Indicators of emotional arousal in mammalian and human vocalizations]. Zhurnal Obschey Biologii. 70(33):210–224. Original (in Russian): Володин, И. А., Володина, Е. В., Гоголева, С. С., & Доронина, Л. О. (2009). Индикаторы эмоционального возбуждения в звуках млекопитающих и человека. Журнал общей биологии, 70(3), 210-224.

White B. 2014. 'Circular breathing'. In: The Oxford Companion to Music. Oxford Music Online. Oxford University Press. [Accessed 2021 Sep 19] https://www.oxfordreference.com/view/10.1093/acref/9780199579037.001.0001/acref-9780199579037-e-1423#acref-9780199579037-e-1423

Wood A. 2020. Social context influences the acoustic properties of laughter. Affective Science. 1 (4):247–256. doi:10.1007/s42761-020-00022-w.