JASA EXPRESS LETTERS | ARTICLE | asa.scitation.org/journal/jel

MRI and acoustic study of angry screaming versus trained belting with vocal health implications

Karen Perta,[1,a] Zhaoyan Zhang,[2,b] Donna Erickson,[3,c] Ryoko Hayashi,[4,d] and Toshiyuki Sadanobu[5,e]

[1]Communication Sciences and Disorders, Elmhurst University, Elmhurst, Illinois 60126, USA
[2]UCLA School of Medicine, UCLA, Los Angeles, California 90095, USA
[3]Haskins Laboratories, New Haven, Connecticut 06511, USA
[4]Graduate School of Intercultural Studies, Kobe University, Kobe, Japan
[5]Graduate School of Letters, Kyoto University, Kyoto, Japan

Abstract: Intense vocalizations, such as screaming and yelling, may yield negative vocal health consequences. The purpose of this study was to investigate physiological and acoustic differences between emotional versus learned high-intensity voice productions (i.e., primitive screaming versus trained belt for healthy yelling). We hypothesized that (1) angry screaming would be produced with constrictive vocal tract and laryngeal actions, similar to swallowing physiology; and (2) trained belting would override natural constrictive tendencies to produce intense voice with reduced phonotraumatic risk. In this study, we present single-subject magnetic resonance imaging and acoustic observations that are consistent with our hypotheses.

[Editor: Douglas D. O'Shaughnessy] https://doi.org/10.1121/10.0044161

Received: 23 February 2026 **Accepted:** 1 June 2026 **Published Online:** 15 June 2026

1. Introduction

In both human and nonhuman primates, the primary functions of the larynx and supralaryngeal vocal tract relate to swallowing and airway protection. Humans and nonhuman primates also use the vocal apparatus for emotional vocalization that serves a critical function in socialization and affective communication.[1] However, humans evolved neural and anatomical mechanisms that make speech and highly specialized voicing possible.[2] It is currently hypothesized that, though the inputs may differ, an integrated sensorimotor neural system controls both emotional and learned vocalization.[3,4] Despite advances in understanding the human neural bases of emotional versus volitional vocalization,[3–6] very little is known about the physiological mechanisms of voice production under extreme, uncontrolled emotional load. Both voice and swallowing involve precise movements of the lips, tongue, jaw, velum, pharynx, and larynx that can be placed under various levels of volitional control.[4] A protective, normal swallow is produced with propulsive tongue base retraction, upward, forward movement of the larynx (i.e., hyolaryngeal excursion), epiglottic inversion, arytenoid adduction/approximation, a closed velopharyngeal port, and pharyngeal constriction.[4,7] Given the primitive nature of both swallowing and emotional angry screaming, this study sought to answer the question, are angry vocalizations produced with constrictive laryngeal and pharyngeal configurations as seen in swallowing physiology?

Emotional, primitive vocalization versus learned, trained voice production is a critical distinction with practical applications. Any high activation, intense voice production—whether emotional or trained—has the potential to cause phonotrauma.[8,9] Understanding the physiological causes of phonotrauma is essential to clinical voice therapy and singing pedagogy in order to teach voice users how to produce the desired sound in a manner that minimizes the risk for phonotrauma and vocal fatigue. For example, in voice therapy, teaching high-intensity voice qualities such as belt (a quality often heard in contemporary singing styles, such as musical theater, pop, and/or rock) for healthy yelling can address the common concern of being heard in noisy environments or over large distances.[10] While there have been many studies on the acoustics and physiology of the belt voice[9,11] and the effect of emotion on voice/speech acoustics and vocal tract shapes,[12–16] the physiologic and acoustic differences between trained belting and primitive angry screaming are still unknown. Although phonotraumatic risks may differ, no prior study has compared primitive screaming with trained belting using magnetic resonance imaging (MRI) and acoustic data.

[a]Corresponding author: karen.perta@elmhurst.edu
[b]Electronic mail: zyzhang@ucla.edu
[c]Electronic mail: EricksonDonna2000@gmail.com
[d]Electronic mail: rhayashi@kobe-u.ac.jp
[e]Electronic mail: sadanobu.toshiyuki.3x@kyoto-u.ac.jp

The purpose of this study was to compare physiological and acoustic features associated with primitive, angry screaming versus trained belt for healthy yelling. We hypothesized that (i) angry vocalizations would involve constricted vocal tract and laryngeal gestures, similar to swallowing physiology, and (ii) trained vocalization would override innate, constrictive tendencies to prevent vocal fatigue and phonotrauma. To test these hypotheses, we used real-time MRI and acoustic measures to analyze emotional, uncontrolled angry screaming versus learned, controlled belting in a single, vocally trained female subject. Our single-subject observations support these hypotheses that compared to trained belt, angry screaming was produced with a more pressed source configuration and constrictive vocal tract actions that contribute to phonotrauma.

2. Materials and methods

2.1 Subject

At the time of scanning, the subject was a 30-year-old female certified in Estill Voice Training at the Estill Master Trainer level with 14 years of singing experience in a variety of styles and a licensed speech-language pathologist specializing in voice and voice disorders. She denied illness or voice problems on the day of scanning and had a history of normal video laryngeal stroboscopy (VLS).

2.2 MRI procedure and task

All MRI data were acquired using a 3 tesla Siemens MAGNETOM Verio scanner (Siemens, Munich, Germany) with a 20-channel head and neck coil. Audio data were captured via optical microphone (FOMRI-I; Optoacoustics Ltd, Or Yehuda, Israel). Imaging was performed in a single midsagittal two-dimensional (2D) gradient-echo real-time sequence at approximately ten frames per second. Sequence parameters derived from DICOM metadata included repetition time (TR) = 3.4 ms, echo time (TE) = 1.38 ms, slice thickness = 10 mm, flip angle = 8°, field of view = $256 \times 256\ mm^2$, and acquisition matrix = 128×84. Images were interpolated to a 256×168 matrix, yielding a reconstructed pixel spacing of $1 \times 1.5\ mm^2$; however, the acquired in-plane resolution, based on field of view and acquisition matrix, was $2.0 \times 3.05\ mm^2$. Accordingly, all linear measurements are interpreted with respect to the acquired spatial resolution ($2.0 \times 3.05\ mm^2$).

The female subject performed a variety of vocal tasks throughout an MRI scanning session of approximately 45 min. Scream was the last task performed. Vocalization tasks were performed with natural breath cycles, resulting in multiple repetitions of each task across the 30-s scan duration. The belt task was performed according to the Estill definition on the /e/ vowel at the subject's most comfortable pitch in a controlled manner.[11] For the emotional vocalization task, the subject was given the prompt, "Please scream like you have never been so angry in your whole life. Do not try to control it." There were no pitch or vowel specifications. The elicited task included intermittent sob-like segments. For each vocal task, MRI images and recorded audio were compiled into video files in .avi format (see Mm. 1).

Mm. 1. MRI video of all vocalizations.

2.3 MRI data analysis

The MRI DICOM files were imported into open-source software program Slicer (3D Slicer 5.8.1; National Alliance for Medical Image Computing, Boston, MA) for visualization, analysis, and measurement.[17] Individual timepoints from the video files for each task were identified for analysis, and the corresponding DICOM file was selected for 2D MRI visualization and measurement. Quantitative MRI analysis was limited to six selected frames sampled from the dynamic acquisition rather than continuous frame-by-frame tracking. During the emotional vocalization task, the subject produced seven screams with emotional sobbing after the fifth scream. As a result, five timepoints from the fifth scream were selected for analysis (Fig. 1), which includes one timepoint pre-scream (PS), three time points during the scream (S1–S3, with subharmonics in S2), and one timepoint during post-scream sob. Since the belt production was relatively stable, only one timepoint was selected for analysis (Fig. 1). Quantitative 2D MRI measures were captured using Slicer's linear and angular measurement tools.

MRI measures were obtained for emotional vocalizations scream and sob as well as trained belt for a total of three voice qualities (Fig. 2). Linear 2D MRI measures, defined similarly to previous work,[18,19] included the following: (1) lip opening, the vertical distance between the lips; (2) jaw opening, the distance between the anterior nasal spine and the most anterior-inferior point of the bone marrow in the corpus mandibulae; (3) jaw protrusion, the length of a line from the most anterior-inferior point of the bone marrow in the corpus mandibulae to the perpendicular intersection of the line used to calculate larynx height; (4) larynx height, the perpendicular distance between the tuberculum anterius atlantis of C2 and the line representing the vocal fold plane; (5) velar height, the perpendicular distance from the tip of the uvula to a reference line representing the hard palate line (i.e., anterior to posterior nasal spine); and (6) pharynx width, the smallest anterior-posterior distance between the back of the tongue and the oropharyngeal wall, superior to the epiglottis. Of note, for larynx height, velar height, and pharynx width, a lower value indicates a higher position or degree of narrowing and vice versa. The angular MRI measure, vocal fold tilt angle, was defined by the vocal fold plane relative to the anterior surface of the cervical spine (Fig. 2). A larger angle reflects a greater degree of tilting.

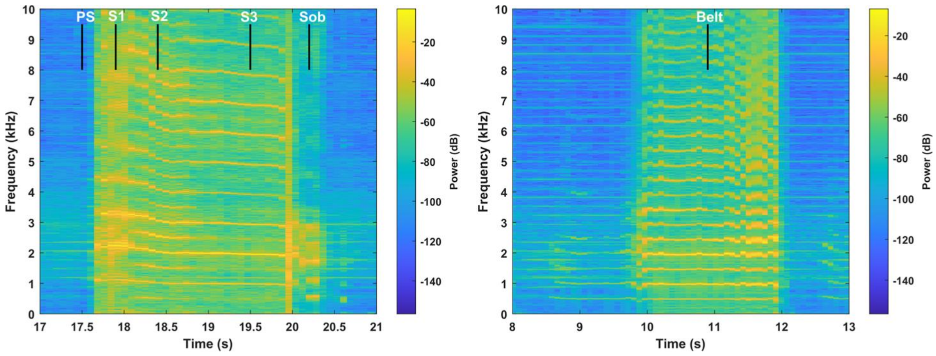

Fig. 1. Spectrograms of emotional vocalization, scream (PS, S1–S3) and sob (left) and trained vocalization, belt (right) obtained from MRI audio. Black lines indicate timepoints for 2D MRI analysis.

2.4 MRI data reliability and analysis

To assess intrarater reliability, the primary author (K.P.) performed two sets of measurements at different timepoints. Both sets of measurements consisted of 42 observations (7 measurements $\times$ 6 slices = 42 observations per set). Absolute mean differences between repeated measurements were 0.3–3.5 mm and 0.7 degrees. Intraclass correlation coefficients (ICCs) (two-way mixed, absolute agreement, and single measures) ranged from 0.958 to 0.999 for all measures, indicating excellent intrarater reliability. Additionally, to calculate interrater reliability, author Z.Z. completed an additional set of 42 measurements. Absolute mean differences between repeated measurements were 0.3–3.5 mm and 1.3 degrees. ICC values ranged from 0.913 to 0.991 for all measures, indicating excellent interrater reliability. Linear and angular 2D measurements were rounded from two decimal places to the nearest whole millimeter or degree. The true acquired spatial resolution (2.0×3.05) defined the practical precision limit of all linear 2D MRI measures. MRI data reported descriptively due to single-subject sample size.

2.5 Acoustic analysis

Audio used for analysis was obtained from the MRI videos after spectral-subtraction scanner noise reduction applied during video construction. Acoustic measures were extracted using the software VoiceSauce with a window size of 25 ms and frame shift of 1 ms.[20] In this study, four measures were reported, including fundamental frequency (f0), $H1^*-H2^*$ (the amplitude difference between the first and second harmonic corrected for vocal tract resonance), $H1^*-H5k^*$ (the

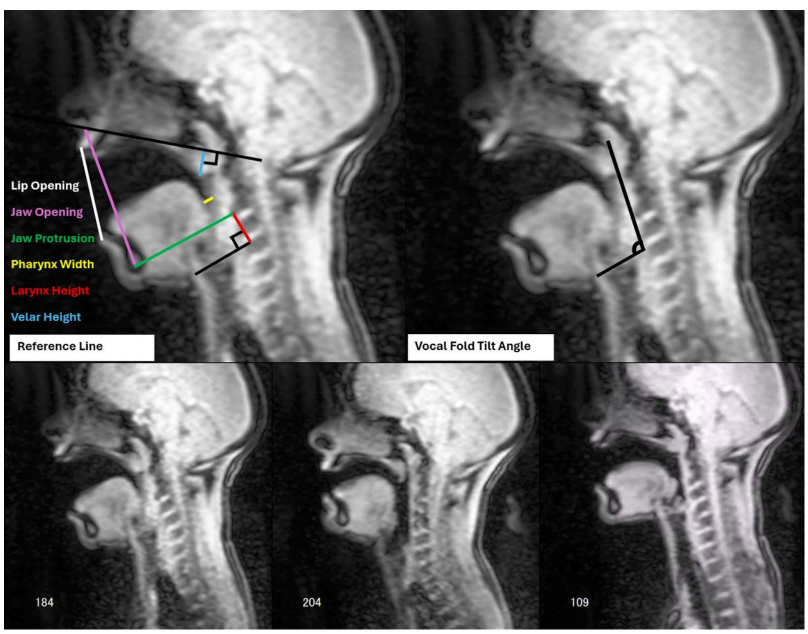

Fig. 2. MRI images representing linear 2D measurements (top left), angular measurement (top right), and the three voice qualities: scream S2 (bottom left), sob (bottom middle), and belt (bottom right).

Table 1. All 2D MRI measures in emotional (scream, sob) versus trained (belt) vocalization.

	PS	Scream 1	Scream 2	Scream 3	Sob	Belt
Lip opening (mm)	40	25	60	57	29	46
Jaw opening (mm)	69	55	91	88	64	79
Jaw protrusion (mm)	81	81	69	71	73	77
Larynx height (mm)	44	30	21	23	46	21
Velar height (mm)	20	15	16	17	20	8
Pharynx width (mm)	16	9	4	5	13	10
Vocal fold tilt angle (°)	99	93	106	106	111	95

amplitude difference between the first harmonic and the harmonic nearest 5 kHz corrected for vocal tract resonance), and cepstral peak prominence (CPP). The timepoints used for acoustic analysis corresponded to the slices used for MRI measurement. Formant frequency data are not reported because the vowel condition was not controlled across tasks.

3. Results

3.1 MRI data

The results of physiological MRI measures for scream, sob, and belt as performed by the primary author (K.P.) are summarized in Table 1. In the angry scream, oral MRI measures across the selected four frames showed a pattern consistent with increased lip opening and jaw opening and decreased jaw protrusion. Pharyngeal measures across four scream timepoints revealed decreased pharynx width as well as increased larynx height, velar elevation with closed velopharyngeal port, and increased vocal fold tilt angle.

In contrast to the scream condition, sob was produced with smaller lip and jaw opening and slightly more jaw protrusion. Sob was also characterized by the lowest velar position with opening of the velopharyngeal port, the lowest vertical larynx position, widest pharynx, and greatest tilting of the vocal fold plane as compared to scream and belt. In the belt condition, lip opening, jaw opening, and pharyngeal narrowing were increased relative to sob but decreased relative to scream. Jaw protrusion and velar elevation were highest during belt. Larynx height was similar between scream and belt, with both qualities characterized by a high vertical larynx position. Vocal fold tilt angle was the smallest in belt, with increased vocal fold tilt angle for scream and sob.

3.2 Acoustic data

Acoustic measures of scream, sob, and belt, including f0, CPP, H1*–H2*, and H1*–H5k*, were calculated from one MRI audio recording per task (Fig. 3). The f0 track was visually verified against the spectrogram.

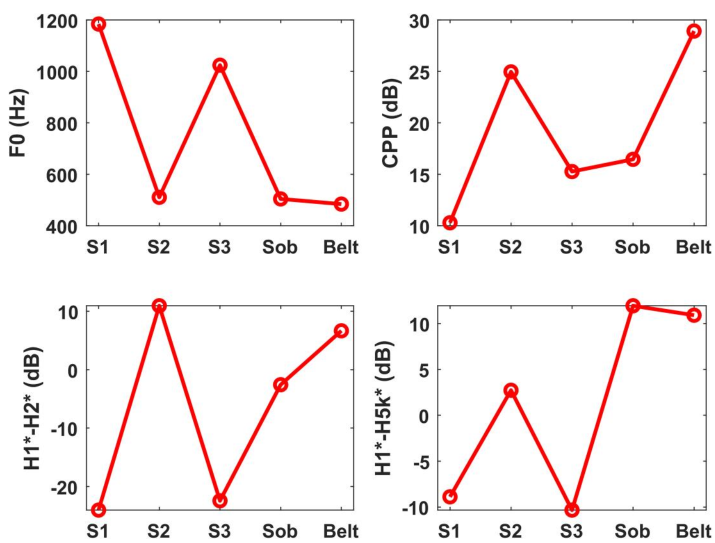

Fig. 3. f0, CPP, H1*–H2*, and H1*–H5k* compared across the three voice qualities. S1–S3 are the three scream conditions in Fig. 1. PS was not included because of no phonation.

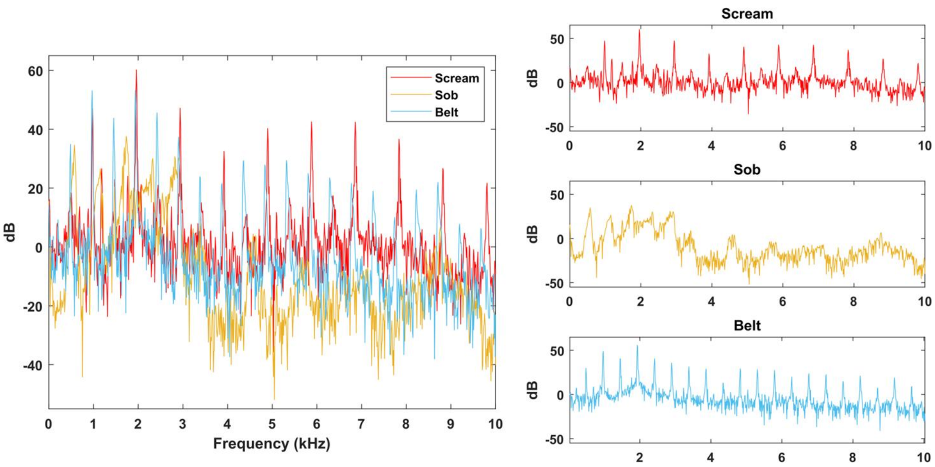

Fig. 4. Spectrum view comparison across emotional (scream, sob) versus trained (belt) voice qualities.

Scream was produced with an f0 over 1000 Hz (except for S2 where f0 was almost half due to subharmonics), almost twice that of sob and belt. CPP is an acoustic measure of irregularity in vocal fold vibration, with low CPP values being associated with low harmonic to noise ratio[21] as well as voice disorders.[22] CPP values were low for emotional vocalizations scream and sob but high in trained belt. Previous simulation studies have shown that $H1^*-H2^*$ and $H1^*-H5k^*$ are primarily controlled by vocal fold vertical thickness as well as vocal fold approximation to a lesser degree, with thicker vocal folds tightly approximated producing lower $H1^*-H2^*$ and $H1^*-H5k^*$.[23,24] Both $H1^*-H2^*$ and $H1^*-H5k^*$ were lowest in scream (except for S2 due to subharmonics) and highest in belt, consistent with a more pressed source configuration in scream. Spectrum view comparison (i.e., scream, sob, belt) is presented in Fig. 4. Intense vocalizations (i.e., scream and belt) are produced with strong intensity in the higher harmonics (4 kHz+), whereas sob demonstrates a characteristic intensity decrease in the upper frequencies.[11]

3.3 Vocal health

After the MRI scans of the emotional screaming and sobbing, the female subject experienced an immediate deterioration in her voice quality. The emotional vocalization scans were the last performed, and the participant did not experience voice difficulty earlier in the session. VLS performed four days after scanning revealed marked edema and hypervascularity as well as small bilateral lesions of the true vocal folds. This acute vocal injury resolved over the course of 3 weeks (see Mm. 2).

Mm. 2. Female subject rigid VLS exams: baseline, 4 days after MRI, 8 days after MRI, 3 weeks after MRI.

4. Discussion

This study provides physiological and acoustic evidence for the differences between primitive, emotional screaming and learned, volitional belting in a single vocally trained female subject. The exaggerated opening of the oral cavity during scream is consistent with prior studies of angry speech.[15,16] While oral cavity size increased in scream, drastic pharyngeal narrowing was evident. High vertical larynx position, closed velopharyngeal port, and increased vocal fold tilt angle were also observed in scream. Though swallowing was not imaged, this combination of posterior movement of the tongue base, elevated velum, and constricted larynx and pharynx resembles the physiology of the normal swallow.[4,7] Acoustic data suggest that these constrictive vocal tract and laryngeal actions were accompanied by a similarly constrictive source configuration, potentially involving increased vocal fold thickness and approximation. Scream in our study had the lowest CPP, $H1^*-H2^*$, and $H1^*-H5k^*$, values consistent with the highest degree of vocal fold medial compression and vertical thickness in vocal fold models.[24,25] Vocal fold compression and increased thickness are known to produce excessively high vocal fold contact pressure during phonation,[26] which may have contributed to the female subject's acute vocal fold injury. To our knowledge, no physiologic or acoustic data from uncontrolled angry screaming currently exist, and prior studies using controlled, sustainable acting and singing tasks[12–16] or extreme vocal effects[27] are not directly comparable to the vocalization analyzed in the present study.

Compared to scream, belt was produced with a similar overall configuration but with decreased mouth opening, a more forward tongue position, and larger pharynx width. In particular, the pharynx was about twice as wide in belt than scream, indicating that the subject was able to override the innate tendency to constrict the vocal tract during belt. The epilaryngeal narrowing observed in belt may have occurred at a degree that reduced vocal fold contact pressure.[28] Although MRI resolution precluded detailed vocal fold imaging, the less constrictive pharynx and more forward tongue position may have been accompanied by a less pressed source configuration, potentially allowing the subject to belt without experiencing noticeable phonotrauma. Higher values in CPP, $H1^*-H2^*$, and $H1^*-H5k^*$ also support the presence of a less constricted source configuration during belt production.

The overall vocal tract configuration for belt in the 2D midsagittal view, including high vertical larynx position, narrowed anterior-posterior pharynx, and forward tongue position with larger oral cavity, aligns with prior findings.[11]

Although the primary emotion behind the vocalization was anger, intermittent moments of uninhibited sobbing followed the screaming. In comparison to scream, sob was produced with the least amount of oral cavity opening, which is consistent with prior studies of sad speech.[16,29] Sobbing was also characterized by an elongated, widened pharynx, open velopharyngeal port, and low laryngeal position, which is opposite that of the normal swallow.[4,7] Reduced constriction and narrowing in the vocal tract during expression of sadness, relative to anger, aligns with prior work.[16] This long, wide vocal tract configuration for emotional sob is also very similar to previous reports of sob quality production during trained singing tasks,[11] but the lowered velum appears unique to emotional sobbing. Compared to scream, emotional sobbing also yielded relatively higher $H1^*-H2^*$ and $H1^*-H5k^*$, suggesting a less pressed source configuration that may decrease phonotraumatic risk.

It is noteworthy that belt and scream were characterized by equally high vertical larynx position despite substantial differences in f0 (i.e., doubled in scream), whereas belt and sob had drastic differences in laryngeal height despite similar f0. Although vocal fold tilt angle was smallest in belt, scream and sob had similarly large vocal fold tilt angle. This large vocal fold tilt angle could have been achieved in different ways for scream versus sob via anterior, downward movement of the thyroid cartilage and/or anterior, superior movement of the cricoid cartilage. Although multiple muscular actions for changes in laryngeal height and laryngeal tilting have been proposed,[30–33] these mechanisms and the relationship between them still are not fully understood and warrant future investigation. Vocal fold configuration and contact pressure may differ based on laryngeal height and/or tilting conditions, which also need to be investigated in future studies.

Both scream and belt were produced with an increased mouth opening and decreased pharynx width. These findings are consistent with prior MRI work that has described belt quality as "megaphone shaped,"[11,34] with large lip/jaw opening and small, narrow pharynx observed in the 2D midsagittal view. In fact, the vocal tract configuration in scream appeared as an exaggerated megaphone shape. Such vocal tract adjustments are often associated with an increased F1, which allows amplification of higher order harmonics and a belt-like voice quality (Fig. 4). On spectrogram visualization (Fig. 1), belt appeared relatively stable. However, scream appeared unstable, exhibiting irregular harmonics at the beginning, going through a brief period of subharmonics, followed by stable harmonics that decrease in frequency and intensity toward the sob. Strong source-filter interaction may have occurred but cannot be quantified without three-dimensional vocal tract geometry.

The single subject sample size is a significant limitation of this study. This work could be further improved by increased spatial resolution and improved MRI image quality, particularly at vocal fold level and in the tracheal region. Also, the MRI analysis was based on selected frames rather than full dynamic quantitative analysis. Belt stability was judged qualitatively, and the single-frame belt comparison is another limitation. Pitch and vowel conditions were not controlled across tasks because authentic emotional vocalization cannot be produced with controlled pitch and vowel, but task discrepancy does complicate the scream versus belt comparison. The MRI environment poses inherent challenges to vocalization tasks, including scanner noise and supine subject positioning. Participant vocal injury may have been compounded by fatigue and cumulative session effects. Future research may aim to include vocal fold imaging during emotional versus trained vocalizations as well as airflow measures.

5. Conclusion

In this single vocally-trained subject, select midsagittal MRI frames and matched acoustic samples suggest similarities between primitive mechanisms of swallowing and screaming. Compared to a controlled belt task, production of emotion-driven screaming appears consistent with larger jaw and lip opening, greater midsagittal pharyngeal narrowing, similar high vertical larynx position, more tilted vocal fold plane, more constrictive vocal fold configuration, lower acoustic periodicity, and higher risk of vocal fold injury. Whether these hypothesis-generating observations can be generalized to a larger number of subjects warrants future investigation.

Acknowledgments

This study was partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grant-in-Aid for Scientific Research (B), 17KT0059. The authors thank Yukiko Noda (ATR-Promotions, Brain Activity Imaging Center, Kyoto, Japan) for technical support related to MRI acquisition parameters.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The study was approved by the ATR Review Board Ethics Committee (17-167), and informed consent was reviewed and signed. The vocal health exams were released with female subject permission (first author K.P.).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1]G. Bryant, "The evolution of human vocal emotion," Emotion Rev. **13**, 25–33 (2021).
[2]P. Lieberman, "Vocal tract anatomy and the neural bases of talking," J. Phonetics **40**, 608–622 (2012).
[3]M. Belyk and S. Brown, "The origins of the vocal brain in humans," Neurosci. Biobehav. Rev. **77**, 177–193 (2017).
[4]C. L. Ludlow, "Central nervous system control of voice and swallowing," J. Clin. Neurophysiol. **32**, 294–303 (2015).
[5]M. Gerbella, C. Pinardi, G. Di Cesare, G. Rizzolatti, and F. Caruana, "Two neural networks for laughter: A tractography study," Cereb. Cortex **31**, 899–916 (2021).
[6]J. Klingbeil, M. Wawrzyniak, A. Stockert, M.-L. Brandt, H.-R. Schneider, M. Metelmann, and D. Saur, "Pathological laughter and crying: Insights from lesion network–symptom mapping," Brain **144**, 3264–3276 (2021).
[7]A. Vose and I. Humbert, "Hidden in plain sight: A descriptive review of laryngeal vestibule closure," Dysphagia **34**, 281–289 (2019).
[8]T. Murray and C. A. Rosen, "Phonotrauma associated with crying," J. Voice **14**, 575–580 (2000).
[9]T. Ikävalko and A.-M. Laukkanen, "Belting by trained singers compared with shouting by untrained nonsingers: How do they differ?," J. Voice (published online 2025).
[10]E. Grillo, J. Wolfberg, K. Perta, J. Van Stan, and K. Steinhauer, "Connecting auditory-perceptual prompts used in voice therapy to anatomy and physiology: Application to the Estill voice model and the rehabilitation treatment specification system," J. Voice (published online 2024).
[11]M. Fleischer, S. Rummel, F. Stritt, J. Fischer, M. Bock, M. Echternach, B. Richter, and L. Traser, "Voice efficiency for difference voice qualities combining experimentally derived sound signals and numerical modeling of the vocal tract," Front. Physiol. **13**, 1081622 (2022).
[12]S. Patel, K. R. Scherer, E. Björkner, and J. Sundberg, "Mapping emotions into acoustic space: The role of voice production," Biol. Psychol. **87**, 93–98 (2011).
[13]J. Sundberg, G. L. Salomão, and K. R. Scherer, "Emotional expressivity in singing: Assessing physiological and acoustic indicators of two opera singers' voice characteristics," J. Acoust. Soc. Am. **155**, 18–28 (2024).
[14]K. R. Scherer, J. Sundberg, B. Fantini, S. Trznadel, and F. Eyben, "The expression of emotion in the singing voice: Acoustic patterns in vocal performance," J. Acoust. Soc. Am. **142**, 1805–1815 (2017).
[15]S. Lee, E. Bresch, J. Adams, A. Kazemzadeh, and S. Narayanan, "A study of emotional speech articulation using a fast magnetic resonance imaging technique," in *Proceedings of Interspeech 2006*, Pittsburgh, PA (2006), pp. 2234–2237.
[16]J. Kim, A. Toutios, S. Lee, and S. S. Narayanan, "Vocal tract shaping of emotional speech," Comput. Speech Lang. **64**, 101100 (2020).
[17]R. Kikinis, S. D. Pieper, and K. G. Vosburgh, "3D slicer: A platform for subject-specific image analysis, visualization, and clinical support," in *Intraoperative Imaging and Image-Guided Therapy* (Springer, New York, 2014), pp. 277–289.
[18]M. Echternach, F. Burk, J. Kirsch, L. Traser, P. Birkholz, M. Burdumy, and B. Richter "Articulatory and acoustic differences between lyric and dramatic singing in Western classical music," J. Acoust. Soc. Am. **155**, 2659–2669 (2024).
[19]K. Perta, Y. Bae, and K. Obert, "A pilot investigation of twang quality using magnetic resonance imaging," Logoped. Phoniatr. Vocol. **46**, 77–85 (2021).
[20]Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu (2011). "VoiceSauce: A program for voice analysis," in *Proceedings of the ICPhS XVII*, pp. 1846–1849.
[21]J. M. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," J. Speech. Lang. Hear. Res. **37**(4), 769–778 (1994).
[22]O. Murton, R. Hillman, and D. Mehta, "Cepstral peak prominence values for clinical voice evaluation," Am. J. Speech. Lang. Pathol. **29**, 1596–1607 (2020).
[23]Z. Zhang, "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," J. Acoust. Soc. Am. **139**(4), 1493–1507 (2016).
[24]Z. Zhang, "Interaction effects in laryngeal and respiratory control of the voice source and vocal fold contact pressure," J. Acoust. Soc. Am. **156**(6), 4326–4335 (2024).
[25]Z. Zhang, "Mechanics of human voice production and control," J. Acoust. Soc. Am. **140**(4), 2614–2635 (2016).
[26]Z. Zhang, "Laryngeal strategies to minimize vocal fold contact pressure and their effect on voice production," J. Acoust. Soc. Am. **148**, 1039–1050 (2020).
[27]M. Aaen, J. McGlashan, N. Christoph, and C. Sadolin, "Extreme vocal effects distortion, growl, grunt, rattle, and creaking as measured by electroglottography and acoustics in 32 health professional singers," J. Voice **38**(3), 795.e21–795.e35 (2024).
[28]Z. Zhang, "Interaction between epilaryngeal and laryngeal adjustments in regulating vocal fold contact pressure," JASA Express Lett. **1**, 025201 (2021).
[29]D. Erickson, K. Yoshida, C. Menezes, A. Fujino, T. Mochida, and Y. Shibuya, "Exploratory study of some acoustic and articulatory characteristics of sad speech," Phonetica **63**, 1–25 (2006).
[30]M. Aaen, N. Christoph, J. McGlashan, and C. Sadolin, "Correlating degree of thyroid tilt independent of f_0 control as a mechanism for phonatory density with egg and acoustic measures across loudness conditions," Folia Phoniatr. Logop. **77**, 319–331 (2025).
[31]Y. T. Hong, K. H. Hong, J. P. Jun, and P. H. Hwang, "The effects of dynamic laryngeal movements on pitch control," Am. J. Otolaryngol. **36**, 660–665 (2015).
[32]E. Vilkman, A. Sonninen, P. Hurme, and P. Körkkö, "External laryngeal frame function in voice production revisited: A review," J. Voice **10**(1), 78–92 (1996).
[33]K. Steinhauer, M. Klimek, and J. Estill, *The Estill Voice Model Theory & Translation* (Estill Voice International, Pittsburgh, PA, 2017).
[34]I. Titze and A. Worley, "Modeling source-filter interaction in belting and high-pitched operatic male singing," J. Acoust. Soc. Am. **126**(3), 1530–1540 (2009).