



Articulation posture influences pitch during singing imagery

Anita Körner¹ · Fritz Strack²

Accepted: 3 May 2023 / Published online: 23 May 2023
© The Author(s) 2023

Abstract

Facial muscle activity contributes to singing and to articulation: in articulation, mouth shape can alter vowel identity; and in singing, facial movement correlates with pitch changes. Here, we examine whether mouth posture causally influences pitch during singing imagery. Based on perception–action theories and embodied cognition theories, we predict that mouth posture influences pitch judgments even when no overt utterances are produced. In two experiments (total $N = 160$), mouth posture was manipulated to resemble the articulation of either /i/ (as in English *meet*; retracted lips) or /o/ (as in French *rose*; protruded lips). Holding this mouth posture, participants were instructed to mentally “sing” given songs (which were all positive in valence) while listening with their inner ear and, afterwards, to assess the pitch of their mental chant. As predicted, compared to the o-posture, the i-posture led to higher pitch in mental singing. Thus, bodily states can shape experiential qualities, such as pitch, during imagery. This extends embodied music cognition and demonstrates a new link between language and music.

Keywords Singing imagery · Pitch · Musical imagery · Language · Embodied cognition

Introduction

The muscles that shape the mouth have several functions, prominently among them articulating speech and expressing emotions. Accordingly, changing mouth posture, specifically, lip corner retraction (vs. lip rounding) has been found to influence mood (Coles et al., 2022) and acoustic properties of articulated vowels (Fagel, 2010). Connecting emotional and articulatory properties, articulation has been found to be associated with word valence, such that words whose vowels require lip retraction were associated with more positive valence than words whose vowels require lip rounding (Körner & Rummer, 2022a). In the present research, we extend the research concerning the influences of mouth posture to pitch, examining whether mouth posture causally influences pitch during mental singing.

Pitch in singing

Vocal sounds are produced by air flow passing from the lungs through the vocal tract. Sound waves are produced in the larynx and then modulated by other articulators before they are emitted. The major determinant of pitch is the vibration speed of the vocal folds, while the configuration of later articulators, for example, tongue and jaw position, mainly influence other aspects of the vocal product, such as vowel formants and breathiness of the voice (e.g., Titze, 1994; Wolfe et al., 2020). Thus, physiologically speaking, mouth posture is not a major determinant of pitch. Accordingly, trained singers are able to vary pitch independent of facial posture (e.g., Lange et al., 2022). Nevertheless, initial evidence suggests that facial movement during singing correlates with pitch production and perception. When singing, facial movement of the singer varies with interval size, so that head movement, eyebrow movement, and lip movement increase with increasing interval size (Thompson & Russo, 2007). Moreover, other people use these cues when predicting sung pitch. Thus, singers’ facial muscle activity influences pitch judgment in listeners (Laeng et al., 2021; Thompson et al., 2010; Thompson & Russo, 2007). The present research directly manipulates facial muscle activity – in a way unrelated to the singing task – and asks singers to evaluate pitch in mental singing.

✉ Anita Körner
anita.koerner@uni-kassel.de

¹ Department of Psychology, University of Kassel,
Holländische Straße 36–38, 34127 Kassel, Germany

² Department of Psychology, University of Würzburg,
Würzburg, Germany

Mental singing refers to an imagery process where one's inner voice (i.e., a covert articulation mechanism) creates a chant that can be perceived with one's inner ear (i.e., an imagery of hearing speech) – without overt utterances (for reviews, see Hubbard, 2010, 2019; Perrone-Bertolotti et al., 2014). Auditory imagery in general is highly accurate, resembling the perception of overt stimuli, for example, in duration, loudness, and pitch (e.g., Farah & Smith, 1983; Halpern, 1988; Wu et al., 2011). Mental singing involves a covert activation of articulation muscles (Smith et al., 1995), leading to increased larynx activation during mental singing (vs. listening to music; Bruder & Wöllner, 2021; see also Pruitt et al., 2019). Moreover, articulation suppression has been found to reduce involuntary song imagery (Beaman et al., 2015) and to interfere with the memory for songs (Wood et al., 2020; cf. Weiss et al., 2021). Thus, covert articulation can be assumed to causally contribute to mental singing, influencing auditory imagery.

Perception–action links

The influence of covert articulation on auditory imagery can be explained by action–perception theories (e.g., Hommel, 2019; Knoblich & Sebanz, 2006; Shin et al., 2010; Witt, 2011). According to action–perception theories, action planning involves predicting the anticipated sensory consequences of the action, so that perception and action involve shared processes and shared representations. Specifically, for speech perception and production, the motor theory of speech perception (Galantucci et al., 2006; Liberman & Mattingly, 1985) postulates that perceiving speech involves the speech motor system. Supporting a causal role of articulation in speech perception, temporary impairment of lip (vs. tongue) motor brain regions has been found to deteriorate, though not completely impair, discrimination of phonemes articulated with the lips (vs. tongue; D'Ausilio et al., 2009; for reviews of evidence for and against motor influences on speech perception, see Galantucci et al., 2006; Skipper et al., 2017).

Similar to speech, for music, there has also been evidence for bidirectional contributions of action and perception processes (e.g., Godøy et al., 2016; Zatorre et al., 2007; for reviews, see Keller, 2012; Maes et al., 2014; Novembre & Keller, 2014; Schiavio et al., 2014). For example, a manipulation of movement rhythm has been found to influence later judgments concerning heard rhythms (Phillips-Silver & Trainor, 2007). Moreover, musical imagery can be disrupted by concurrent auditory or articulatory tasks, so that participants made more mistakes when comparing a heard musical theme to a previously seen musical notation if they concurrently performed (vs. did not perform) a rhythmic or articulatory task (Brodsky et al., 2008; see

also Brown & Palmer, 2013; for other motor influences, see Connell et al., 2013; Jakubowski et al., 2015). In sum, action has been argued to be involved in the perception and understanding of both speech and music.

Related predictions are made by embodied cognition theories. Embodied cognition postulates that information processing involves the simulation of sensory, motor, or emotional experiences (Barsalou, 2008; Kiefer & Pulvermüller, 2012; Winkielman et al., 2018; for reviews, see Glenberg, 2010; Körner et al., 2015; Körner et al., 2023; Meteyard et al., 2012). Interference with these sensorimotor simulations has been found to alter cognitive processing. Concerning the influence of facial posture, previous research concentrated on emotional content (Niedenthal, 2007). Specifically, emotional information was processed faster (Havas et al., 2007) and detected more accurately (Niedenthal et al., 2001, 2009) when participants' facial expression could be congruent (e.g., smiling for positive information) compared to when it had to be incongruent (e.g., smiling for negative information) with the valence of the information. Moreover, participants' mouth posture has been found to influence evaluations (Strack et al., 1988). When muscles involved in smiling were activated by holding a pen with their teeth (vs. protruding lips), participants rated cartoons to be funnier (Strack et al., 1988; for a meta-analysis, see Coles et al., 2019). Note, however, that the boundary conditions of this effect are still under debate (Wagenmakers et al., 2016; see also, Coles et al., 2022; Noah et al., 2018; Strack, 2016). From embodied cognition research, we predict that altering the posture of muscles involved in singing could influence the chant produced during singing imagery.

The influence of mouth posture on pitch

Mouth posture can influence pitch via different pathways: first, as a direct link; second, through associations with fundamental and formant frequencies of vowels; third, through associations with valence. The direct influence of mouth posture on pitch could result from changed vocal tract length. Retraction of the lip corners shortens the vocal tract (Shor, 1978); conversely, rounding and protruding the lips lengthens the vocal tract (Dusan, 2007; Fant, 1960), which facilitates lower-pitched utterances (Ohala, 1984). Accordingly, in chimpanzees, greater (vs. lesser) lip retraction has been found to be associated with higher pitched vocal utterances (Bauer, 1987). Similarly, when human participants were asked to smile while speaking, their pitch increased (Barthel & Quené, 2015; Tartter, 1980; for the reverse influence, see Huron et al., 2009).

Second, facial muscle activity also influences the articulation of vowels; vowels, in turn, differ in their fundamental frequency (which is perceived as pitch) and formant

frequencies.¹ For example, /i/ – the vowel with the greatest lip corner retraction – has a higher fundamental as well as second formant frequency than /o/ – a vowel with strong lip protrusion (Hoole & Mooshammer, 2002). Second formant frequency also influences subjective tone height, so that when the fundamental frequency is identical, vowels with low (vs. high) second formant are perceived as higher (Fowler & Brown, 1997; see also Russo et al., 2019). Consistent with both the relative fundamental and second formant frequencies of /i/ and /o/, yodelers (which consist largely of nonsense syllables) typically employ /o/ for low-pitched passages and /i/ for high-pitched passages (Fenk-Oczlon & Fenk, 2009). Thus, lip retraction (protrusion), by being associated with intrinsically high-pitched (low-pitched) vowels, could increase (decrease) pitch.

Third, the posture–pitch association might result from a common association with valence. High pitch and positive valence have been found to be associated on a conceptual level (Eitan & Timmers, 2010). Moreover, in various species, including humans, high pitch signals friendly, affiliative behavior, whereas low pitch signals aggressive, dominating behavior (Hodges-Simeon et al., 2010; Morton, 1977; Wood et al., 2017). Retracting (vs. protruding) lips might have been used to influence pitch to signal deference (vs. aggression; Ohala, 1984). Relatedly, vowels differ in their perceived valence. Specifically, /i/ has been found to be associated with positive valence and /o/ with negative valence (Rummer et al., 2014; see also Garrido & Godinho, 2021; Körner & Rummer, 2022b; Rummer & Schweppe, 2019; Yu et al., 2021). Moreover, this association between vowels and valence has been found to rest on articulation muscle activity (Körner & Rummer, 2022a). Thus, lip retraction (protrusion) could further increase pitch because positive (negative) valence is associated with high (low) pitch.

In sum, several processes might cause facial postures with retracted (vs. protruded) lips to be associated with high (vs. low) pitched vocalizations. In the present research, we manipulated participants' facial posture and assessed pitch during singing imagery. Singing involves pitch changes relating to the melody as well as to the articulation of the lyrics. The vowels in lyrics differ in fundamental and formant frequencies. As described above, during singing, facial posture could alter pitch, either by itself or through its relation to vowel frequencies or vowel valence. Because covert articulation is instrumental in singing imagery, we hypothesized that facial posture would influence pitch even though participants only imagined singing.

¹ Formants are energy peaks in the frequency spectrum. In speech, formants distinguish phonemes; the first and second formants are frequently sufficient to distinguish vowels.

Experiment 1

Experiment 1 provides a first test regarding whether mouth posture influences pitch during singing imagery. Participants were asked to adopt a posture associated with the pronunciation of either /i/ or /o/ while singing mentally. We hypothesized that an i-posture compared to an o-posture would lead to higher pitch during singing imagery.

Method

Transparency and openness

We report sample size determination, all data exclusions (if any), all manipulations, and all measures. All data, analysis code, and research materials are available via the Open Science Framework at <https://osf.io/uzfq8/>. Both experiments (hypothesis, design, and main analysis) were pre-registered; Experiment 1: <https://osf.io/g96q4/>; Experiment 2: <https://osf.io/q76hu>.

Participants

We used $d = 0.40$ as the smallest effect size of interest. To detect $d = 0.40$ with $1 - \beta = .90$ and $\alpha = .05$ in a one-tailed between-participants t -test, 216 participants are required. We pre-registered sequential testing (Lakens, 2014) with two interim analyses. Calculating the adjusted alpha-levels using *GroupSeq* (version 1.3.4; Pahl, 2018) with α^*t^1 as spending function resulted in the following p -values for rejecting the null-hypothesis: $p < .017$ with 72 participants, $p < .023$ with 144 participants, and $p < .030$ with 216 participants. The first interim analysis satisfied the stopping criterion, resulting in 72 participants (49 female, 23 male; 58 German native speakers; $M_{age} = 26$ years, $SD_{age} = 6$ years, recruited on a German university campus, without screening for musical training).

Mental singing task

Participants were informed that their task was to imagine singing parts of given songs; singing “in their heads” as clearly as possible while listening with their “inner ear,” but taking care not to sing aloud. For each song excerpt, participants first read the title. When they were ready to begin singing, they were shown the sheet music for the song excerpt with the lyrics underneath the music staff lines. Participants indicated both the beginning and end of their mental singing by key presses. After mentally singing the provided song excerpt, they evaluated the song and their mental singing on five dimensions (e.g., speed and clarity). Embedded was the dependent measure. Participants answered (translated) *How high did the song sound in your head (we refer to pitch)?*

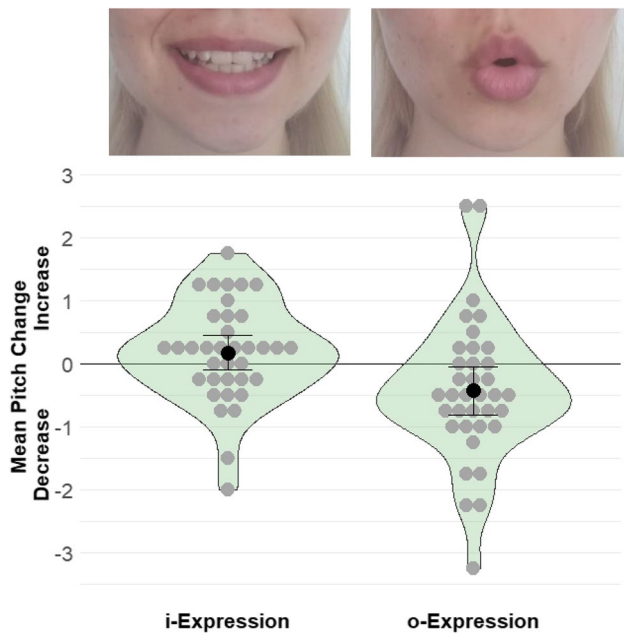


Fig. 1 Pictures explaining the instructions for the facial posture (**upper panel**) and plot of the results (**lower panel**) of Experiment 1. The results show how pitch judgments changed from the baseline to the experimental block depending on mouth posture. *Note.* The black dots and error bars depict means and 95% confidence intervals, the gray dots depict individual participant means, and the shapes are density plots

on a 7-point scale ranging from 1 (*very deep*) to 7 (*very high*). When all evaluations were completed, the next song followed.

There were two sets of songs, each consisting of one Christmas carol, two classic pop songs and one children's song. Participants received one set during the baseline phase and the other during the experimental phase (counterbalanced across participants). Valence evaluations by an independent group of participants indicated that these songs were on average experienced to be mildly to moderately positive (see the Appendix in Online Supplemental Material (OSM)).

Procedure

Participants first provided informed consent. Then a baseline phase ensued, consisting of four singing imagery trials. For this, participants were asked to sing mentally without receiving instructions about facial posture. Each singing trial consisted of the mental singing phase, followed by the evaluation phase.

² As German has a close grapheme to phoneme mapping, the letter graphemes indicated which phonemes were meant.

Then the experimental phase ensued. Participants were randomly assigned to either the i-posture or the o-posture. They were asked to adopt a facial expression as if they were articulating the letter i (or o)² for another four songs. To illustrate the mouth posture, participants were shown the lower part of a face adopting the desired pose; see Fig. 1 upper panel. Before each trial in the experimental phase, participants were asked to adopt the pose and maintain it during the mental singing. For the subsequent evaluations, participants could relax their face.

After they had mentally sung and evaluated all eight songs, participants were asked to provide demographic information and to guess the purpose of the manipulation.

Results and discussion

The dependent measure was the mean pitch judgment in the experimental trials minus mean pitch judgment in the baseline trials. Accordingly, pitch values greater (lesser) than 0 indicate an increase (decrease) in pitch relative to baseline.

Confirming our hypothesis, pitch judgments were higher ($M = 0.18$, $SD = 0.79$) for participants who adopted an i-posture than for participants who adopted an o-posture ($M = -0.42$, $SD = 1.14$), $t(62) = 2.62$, $p = .006$ (one-tailed), $d = 0.62$, 95% CI [0.14; 1.09], see Fig. 1.³ Exploratory analyses indicated that, relative to the baseline, adopting an o-posture decreased pitch judgments, $t(35) = 2.23$, $p = .032$, $d_z = 0.37$, 95% CI [0.03; 0.71], while adopting an i-posture did not significantly increase pitch judgments, $t(35) = 1.38$, $p = .178$, $d_z = 0.23$, 95% CI [-0.10; 0.56].

Thus, Experiment 1 provides first evidence that articulation posture influences the pitch of singing imagery, with increased pitch for an i-posture compared to an o-posture. However, it cannot be excluded that this effect was caused by participants' thinking about the articulation of /i/ or /o/. To preclude this possibility, the manipulation in Experiment 2 did not refer to vowel articulation.

Experiment 2

Experiment 2 employed a slightly different manipulation of articulation posture – participants held a straw in their mouth similar to Strack et al. (1988). We expected to replicate the results of Experiment 1 even when participants were not led to think of their facial posture in terms of vowel articulation.

³ An alternative analysis, incorporating random effects for both participants and stimuli and treating responses as ordinal responses generated by a latent continuous scale, yields similar results, see the OSM.

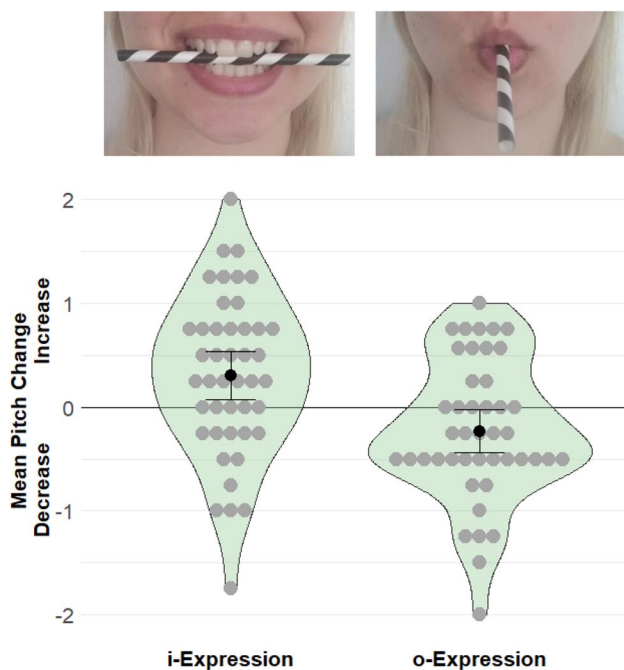


Fig. 2 Pictures explaining the instructions for the facial posture (**upper panel**) and plot depicting the results (**lower panel**) of Experiment 2. The results show how pitch judgments changed from the baseline to the experimental block depending on mouth posture. *Note.* The black dots and whiskers depict means and confidence intervals, the gray dots depict individual participant means, and the shapes are density plots

Method

Participants

We again used sequential testing, this time with one intermediate test after about 55% of the maximum target N of 150 (resulting from a power analysis with $d = 0.41^4$, $1 - \beta = .80$, and $\alpha = .05$ for the one-tailed between-participants t -test). The adjusted alpha-levels were $p < .028$ with about 85 participants and $p < .034$ with 150 participants. This resulted in 88 participants (56 female, 32 male; 83 German native speakers; $M_{\text{age}} = 29$ years, $SD_{\text{age}} = 10$ years; recruited from a local participant pool of a German university that mainly consists of undergraduates; they were not screened for musical training).

Procedure

Materials and procedure were identical to Experiment 1 except for minor wording changes and the facial

manipulation. After the baseline trials, participants were asked to hold a straw in their mouth to control facial movements during mental singing. Half the participants were instructed to hold the straw vertically using their teeth while avoiding their lips touching the straw. The remaining participants were instructed to hold the straw with their protruded lips while avoiding their teeth touching the straw. To ensure correct execution, the corresponding picture of Fig. 2, upper panel, was shown. For the evaluation phase, participants were asked to remove the straw, and afterwards they were reminded to replace it between their teeth/lips.

Results and discussion

As in Experiment 1, the pitch judgments were significantly higher for participants who adopted an i-posture ($M = 0.31$, $SD = 0.77$) than for participants who adopted an o-posture ($M = -0.23$, $SD = 0.67$), $t(84) = 3.51$, $p < .001$ (one-tailed), $d = 0.75$, 95% CI [0.31; 1.18], see Fig. 2.⁵ Exploratory analyses indicated that, compared to baseline, adopting an o-posture decreased pitch judgments, $t(43) = 2.31$, $p = .026$, $d_z = 0.35$, 95% CI [0.04; 0.65], while adopting an i-posture increased pitch judgments, $t(43) = 2.64$, $p = .012$, $d_z = 0.40$, 95% CI [0.09; 0.70].

Thus, Experiment 2 replicates the results of Experiment 1. A posture resembling the articulation of /i/ resulted in higher pitch judgments than a posture resembling the articulation of /o/.

General discussion

The present work examined whether mouth posture alters pitch during singing imagery. In two experiments, we found that relative to baseline, participants who retracted their lips in a manner similar to the articulation of /i/ experienced higher pitch during singing imagery and participants who protruded their lips in a manner similar to the articulation of /o/ experienced lower pitch during singing imagery. This influence occurred both when participants were instructed to shape their mouth as if to articulate /i/ (vs. /o/; Experiment 1) and when their mouth posture was manipulated by holding a straw (Experiment 2). Thus, mouth posture influenced pitch in singing imagery.

In the present research, our sample was not selected for musical proficiency. Musical training has been found to improve musical imagery ability, especially concerning pitch (Aleman et al., 2000; Janata & Paroo, 2006). Therefore, we

⁴ As sequential testing can lead to overestimated effect sizes (Lakens, 2014), we used two-thirds of the effect size observed in Experiment 1.

⁵ As in Experiment 1, an ordinal mixed effects analysis yields similar results, see OSM. Additionally, the OSM also includes a correlation analysis indicating that increasingly positive song valence might be associated with increasing effect sizes.

would expect even clearer and more accurate pitch imagery for musical experts, so that the effect of mouth posture on pitch could be stronger for musicians than for non-musicians. However, trained singers need to be able to independently modulate mouth posture for articulation and pitch. Accordingly, for singers, mouth posture and pitch might be decoupled, so that the effect of mouth posture on pitch might be reduced. Whether this is indeed the case needs to be determined in future research.

Another open question is whether mouth posture influences pitch *production* (i.e., the inner voice) or pitch *perception* (i.e., the inner ear). If, in accordance with the above presented potential mechanisms (oral tract length, vowel association, or valence association) mouth posture influences the inner voice, then overt recreation of the acoustic frequency produced by the inner voice should differ depending on mouth posture. However, if mouth posture influences the inner ear, then the objective frequencies of the mental singing might be identical; but hearing the same frequency would lead to differing sensations of pitch depending on mouth posture (for a related finding, see Hostetter et al., 2019). Future research needs to determine whether mouth posture distorts the inner voice or the inner ear.

Strictly speaking, we cannot rule out a third possibility, namely that mouth posture influences pitch only retrospectively. However, the repeated nature of the task renders this possibility unlikely. After a few trials of the baseline block, participants probably remember the questions and form their judgments already during the singing imagery. Moreover, in the experimental block, participants were asked to relax their face during judgments. Thus, any influence during the judgment phase should have weakened the influence of mouth posture because mouth posture did not differ anymore between the conditions. Nevertheless, future research may rule out the contribution of post-singing processes on pitch judgments by asking participants to evaluate pitch during the singing period.

A possible moderator for the present finding is song valence. Facial muscle activity is involved not only in articulation and pitch production but also in emotional processes. An exploratory correlation between song valence and effect size suggests that songs that are more positive might have larger effect sizes; however, this correlation was not significant (see OSM). In the present research, all employed songs were positive. Future research needs to examine whether the influence of facial posture on pitch also holds for negative songs or conversely only obtains for positive songs.

By finding that mouth posture influences pitch during singing imagery, the present research contributes to the literature on language–music similarities (Nayak et al., 2022; Patel, 2010). Although it is clear that the cognitive processing of speech and music is not identical (e.g., Ilie

& Thompson, 2006; Zatorre et al., 2002), various parallels and mutual influences have been observed (Canette et al., 2020; Coffey et al., 2017; Deutsch et al., 2011; Gordon et al., 2015; Slevc et al., 2009; Thompson et al., 2004; Weiss & Trehub, 2022; Yang et al., 2022). Pitch changes in music and language, for example, have been found to be processed by overlapping brain regions (Schön et al., 2010) and to share mechanisms, so that, for example, speaking a tone language is associated with enhanced musical pitch perception (Bidelman et al., 2013; Giuliano et al., 2011; Pfordresher & Brown, 2009). The present research also speaks to a connection between language and music by demonstrating that adopting a mouth posture for vowel articulation influences pitch perception in music imagery. Thus, complementing previous research that mainly examined individual differences or long-term training effects, we find that a situational manipulation of the articulation posture influences the experiential nature of music perception.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-023-02306-1>.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

Data availability All data and materials are available at <https://osf.io/uzfq8>.

Code availability All code is available at <https://osf.io/uzfq8>.

Declarations

Conflicts of interest/Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval This research was conducted in line with the principles of the Declaration of Helsinki. Our local ethics committee only examines studies that entail heightened participant risks, which was not the case in the present research.

Consent to participate In both Experiments, participants provided informed consent.

Consent for publication In both Experiments, participants provided consent for publication of their anonymous data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aleman, A., Nieuwenstein, M. R., Böcker, K. B., & de Haan, E. H. (2000). Music training and mental imagery ability. *Neuropsychologia*, 38(12), 1664–1668.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Barthel, H., & Quené, H. (2015). Acoustic-phonetic properties of smiling revised: Measurements on a natural video corpus. In *In proceedings of the 18th international congress of phonetic sciences*. The University of Glasgow.
- Bauer, H. R. (1987). Frequency code: Orofacial correlates of fundamental frequency. *Phonetica*, 44(3), 173–191.
- Beaman, C. P., Powell, K., & Rapley, E. (2015). Want to block earworms from conscious awareness? B(u)y gum! *Quarterly Journal of Experimental Psychology*, 68(6), 1049–1057.
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One*, 8(4), Article e60676.
- Brodsky, W., Kessler, Y., Rubinstein, B. S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 427–445.
- Brown, R. M., & Palmer, C. (2013). Auditory and motor imagery modulate learning in music performance. *Frontiers in Human Neuroscience*, 7, Article 320.
- Bruder, C., & Wöllner, C. (2021). Subvocalization in singers: Laryngoscopy and surface EMG effects when imagining and listening to song and text. *Psychology of Music*, 49(3), 567–580.
- Canette, L. H., Lalitte, P., Bedoin, N., Pineau, M., Bigand, E., & Tillmann, B. (2020). Rhythmic and textural musical sequences differently influence syntax and semantic processing in children. *Journal of Experimental Child Psychology*, 191, Article 104711.
- Coffey, E. B., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69.
- Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin*, 145(6), 610–651.
- Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L., Willis, M. L., Foroni, F., Reggev, N., Mokady, A., Forscher, P. S., Hunter, J. F., Kaminski, G., Yüryük, E., Kapucu, A., Nagy, T., Haidu, N., Tejada, J., Freitag, R. M. K., et al. (2022). A multi-lab test of the facial feedback hypothesis by the many smiles collaboration. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-022-01458-9>
- Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, 81(1), 124–130. <https://doi.org/10.1016/j.bandc.2012.09.005>
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381–385.
- Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, 129(4), 2245–2252.
- Dusan, S. (2007). Vocal tract length during speech production. In *Proceeding of the Eighth Annual Conference of the International Speech Communication Association*, 1366–1369.
- Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114(3), 405–422.
- Fagel, S. (2010). Effects of smiling on articulation: Lips, larynx and acoustics. In A. Esposito, N. Campbell, C. Vogel, A. Hussain, & A. Nijholt (Eds.), *Development of multimodal interfaces: Active listening and synchrony* (Lecture notes in computer science) (Vol. 5967). Springer. https://doi.org/10.1007/978-3-642-12397-9_25
- Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics*, 33(5), 475–478.
- Fant, G. (1960). *Acoustic theory of speech production*. Mouton.
- Fenk-Oczlon, G., & Fenk, A. (2009). Some parallels between language and music from a cognitive and evolutionary perspective. *Musicae Scientiae*, 13(2_suppl), 201–226.
- Fowler, C. A., & Brown, J. M. (1997). Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, 59(5), 729–738.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377.
- Garrido, M. V., & Godinho, S. (2021). When vowels make us smile: The influence of articulatory feedback in judgments of warmth and competence. *Cognition and Emotion*, 35(5), 837–843.
- Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, 2, Article 146.
- Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4), 586–596.
- Godøy, R. I., Song, M., Nymo, K., Haugen, M. R., & Jensenius, A. R. (2016). Exploring sound-motion similarity in musical experience. *Journal of New Music Research*, 45(3), 210–222.
- Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in Psychology*, 6, Article 1777.
- Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 434–443.
- Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, 14(3), 436–441.
- Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, 21(4), 406–427.
- Hommel, B. (2019). Theory of event coding (TEC) V2.0: Representing and controlling perception and action. *Attention, Perception, & Psychophysics*, 81(7), 2139–2154.
- Hoole, P., & Mooshammer, C. (2002). Articulatory analysis of the German vowel system. In P. Auer (Ed.), *Silbenschnitt und Tonakzent* (pp. 129–159). <https://doi.org/10.1515/9783110916447.129>
- Hostetter, A. B., Dandar, C. M., Shimko, G., & Grogan, C. (2019). Reaching for the high note: Judgments of auditory pitch are affected by kinesthetic position. *Cognitive Processing*, 20(4), 495–506.
- Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, 136(2), 302–329.
- Hubbard, T. L. (2019). Neural mechanisms of musical imagery. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 521–545). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198804123.013.21>
- Huron, D., Dahl, S., & Johnson, R. (2009). Facial expression and vocal pitch height: Evidence of an intermodal association. *Empirical Musicology Review*, 4(3), 93–100.
- Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23(4), 319–330.

- Jakubowski, K., Halpern, A. R., Grierson, M., & Stewart, L. (2015). The effect of exercise-induced arousal on chosen tempi for familiar melodies. *Psychonomic Bulletin & Review*, 22(2), 559–565.
- Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception & Psychophysics*, 68(5), 829–844.
- Keller, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, 1252(1), 206–213.
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805–825.
- Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, 15(3), 99–104.
- Körner, A., Castillo, M., Drijvers, L., Fischer, M. H., Günther, F., Marelli, M., Platonova, O., Rinaldi, L., Shaki, S., Trujillo, J. P., Tsaregorodtseva, O., & Glenberg, A. M. (2023). Embodied processing at six linguistic granularity levels: A consensus paper. *Journal of Cognition*. Advance online publication. <https://doi.org/10.5334/joc.231>
- Körner, A., Topolinski, S., & Strack, F. (2015). Routes to embodiment. *Frontiers in Psychology*, 6, Article 940.
- Körner, A., & Rummer, R. (2022). Articulation contributes to valence sound symbolism. *Journal of Experimental Psychology: General*, 151(5), 1107–1114.
- Körner, A., & Rummer, R. (2023). Valence sound symbolism across language families: A comparison between Japanese and German. *Language and Cognition*, 15(2), 337–354. <https://doi.org/10.1017/langcog.2022.39>
- Laeng, B., Kuyateh, S., & Kelkar, T. (2021). Substituting facial movements in singers changes the sounds of musical intervals. *Scientific Reports*, 11, Article 22442.
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, 44(7), 701–710.
- Lange, E. B., Fänderich, J., & Grimm, H. (2022). Multisensory integration of musical emotion perception in singing. *Psychological Research*, 86, 2099–2114.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.
- Maes, P. J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, (4), Article 1008. <https://doi.org/10.3389/fpsyg.2013.01008>
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7), 788–804.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855–869.
- Nayak, S., Coleman, P. L., Ladányi, E., Nitin, R., Gustavson, D. E., Fisher, S., Magne, C. L., & Gordon, R. L. (2022). The musical abilities, pleiotropy, language, and environment (MAPLE) framework for understanding musicality-language links across the lifespan. *Neurobiology of Language*, 3(4), 615–664. https://doi.org/10.1162/nol_a_00079
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316(5827), 1002–1005.
- Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion*, 15(6), 853–864.
- Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, 96(6), 1120–1136.
- Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology*, 114(5), 657–664.
- Novembre, G., & Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, 8, 603.
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F₀ of voice. *Phonetica*, 41(1), 1–16.
- Pahl, R. (2018). GroupSeq: A GUI-based program to compute probabilities regarding group sequential designs. *The Comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/>
- Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciú, M., & Loevenbrück, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220–239.
- Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, 71(6), 1385–1398.
- Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, 105(3), 533–546. <https://doi.org/10.1016/j.cognition.2006.11.006>
- Pruitt, T. A., Halpern, A. R., & Pfordresher, P. Q. (2019). Covert singing in anticipatory auditory imagery. *Psychophysiology*, 56(3), e13297.
- Rummer, R., & Schweppe, J. (2019). Talking emotions: Vowel selection in fictional names depends on the emotional valence of the to-be-named faces and objects. *Cognition and Emotion*, 33(3), 404–416.
- Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion*, 14(2), 246–250.
- Russo, F. A., Vuvan, D. T., & Thompson, W. F. (2019). Vowel content influences relative pitch perception in vocal melodies. *Music Perception: An Interdisciplinary Journal*, 37(1), 57–65.
- Schiavio, A., Menin, D., & Matyja, J. (2014). Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology: Music, Mind, and Brain*, 24(4), 340–343.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J. L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage*, 51(1), 450–461.
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136(6), 943–974.
- Shor, R. E. (1978). The production and judgment of smile magnitude. *The Journal of General Psychology*, 98(1), 79–96.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105.
- Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374–381.
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33(11), 1433–1454.
- Strack, F. (2016). Reflection on the smiling registered replication report. *Perspectives on Psychological Science*, 11(6), 929–930.
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54(5), 768–777.
- Tartrot, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27(1), 24–27.
- Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, 18(9), 756–757.

- Thompson, W. F., Russo, F. A., & Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychonomic Bulletin & Review*, 17(3), 317–322.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4(1), 46–64.
- Titze, I. R. (1994). *Principles of voice production*. Prentice-Hall.
- Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Jr., & Bulnes, L. C. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, 11(6), 917–928.
- Weiss, M. W., Bissonnette, A. M., & Peretz, I. (2021). The singing voice is special: Persistence of superior memory for vocal melodies despite vocal-motor distractions. *Cognition*, 213, 104514.
- Weiss, M. W., & Trehub, S. E. (2023). Detection of pitch errors in well-known songs. *Psychology of Music*, 51(1) 172–187. <https://doi.org/10.1177/03057356221087447>
- Winkielman, P., Coulson, S., & Niedenthal, P. (2018). Dynamic grounding of emotion concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170127.
- Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, 20(3), 201–206.
- Wolfe, J., Garnier, M., Henrich Bernardoni, N., & Smith, J. (2020). The mechanics and acoustics of the singing voice. In F. A. Russo, B. Ilari, & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing* (pp. 64–78). Routledge. <https://doi.org/10.4324/9781315163734-5>
- Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PLoS ONE*, 12(8), e0183811.
- Wood, E. A., Rovetti, J., & Russo, F. A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition*, 145, 105622.
- Wu, J., Yu, Z., Mai, X., Wei, J., & Luo, Y. (2011). Pitch and loudness information encoded in auditory imagery as revealed by event-related potentials. *Psychophysiology*, 48(3), 415–419.
- Yang, X., Shen, X., Zhang, Q., Wang, C., Zhou, L., & Chen, Y. (2022). Music training is associated with better clause segmentation during spoken language processing. *Psychonomic Bulletin & Review*, 29, 1472–1479. <https://doi.org/10.3758/s13423-022-02076-2>
- Yu, C. S.-P., McBeath, M. K., & Glenberg, A. M. (2021). The gleam-glum effect: /i:/ versus /ʌ/ phonemes generically carry emotional valence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(7), 1173–1185. <https://doi.org/10.1037/xlm0001017>
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.