DOI: 10.1111/nyas.14806

ORIGINAL ARTICLE

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement

Lara Pearson[1] | **Wim Pouw**[2,3]

[1]Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

[2]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands

[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Correspondence
Lara Pearson, Department of Music, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, Frankfurt am Main, Hessen 60322, Germany.
Email: lara.pearson@ae.mpg.de

Lara Pearson and Wim Pouw shared first authorship.

Abstract

In many musical styles, vocalists manually gesture while they sing. Coupling between gesture kinematics and vocalization has been examined in speech contexts, but it is an open question how these couple in music making. We examine this in a corpus of South Indian, Karnatak vocal music that includes motion-capture data. Through peak magnitude analysis (linear mixed regression) and continuous time-series analyses (generalized additive modeling), we assessed whether vocal trajectories around peaks in vertical velocity, speed, or acceleration were coupling with changes in vocal acoustics (namely, F0 and amplitude). Kinematic coupling was stronger for F0 change versus amplitude, pointing to F0's musical significance. Acceleration was the most predictive for F0 change and had the most reliable magnitude coupling, showing a one-third power relation. That acceleration, rather than other kinematics, is maximally predictive for vocalization is interesting because acceleration entails force transfers onto the body. As a theoretical contribution, we argue that gesturing in musical contexts should be understood in relation to the physical connections between gesturing and vocal production that are brought into harmony with the vocalists' (enculturated) performance goals. Gesture–vocal coupling should, therefore, be viewed as a neuro–bodily distributed aesthetic entanglement.

KEYWORDS
cross-modal correspondences, gesture–speech physics, gesture–vocal coupling, South Indian music, vocal music

INTRODUCTION

Across a wide range of musical styles worldwide, vocalists tend to gesture manually while they sing. In existing research, such co-singing gesture practices have been analyzed with regard to communication, expressivity, transmission, iconicity/metaphor, and perceived effort, often as part of broader discussions of music embodiment.[1–5] However, fundamental questions remain unanswered regarding the coupling between gesture and sound—namely, what features of vocal sound and gesture kinematics are most closely coupled, and in what way. In this study, we address these questions for the insight they can provide into why performers gesture as they do. Alongside this empirical study, we make a theoretical contribution proposing that the physical connection between the sound-producing and movement-producing systems should be taken into account alongside cognitive and cultural considerations, in order to better understand multimodality in human expressivity.

We address these issues in the context of a South Indian musical practice, Karnatak vocal music, chosen for the following reasons. In both Karnatak music and related North Indian styles, vocalists tend

to gesture spontaneously while performing, and as a result, there is already a small body of research on gesturing in these practices. This study is thus part of a larger inquiry into gesture and vocal performance in Indian musical contexts. Meanwhile, as the wider field of music and gesture/movement studies is still heavily skewed toward a focus on Western Art Music, jazz, and popular music, this study contributes to increasing diversity among the styles examined. We propose that studies of real musical practices across diverse cultural contexts can play an important role in understanding connections between gesture and sound production in performance contexts.

Karnatak vocalists frequently gesture while singing, producing a variety of tracing, pointing, flicking, pushing, pulling, and stretching motions (for an example, see https://youtu.be/INk1KvYOf8U). These hand and upper body gestures do not comprise a formal system of symbols and referents; instead, performers experience their gesturing as being spontaneous.[4] Nevertheless, similarities can be found between the gestures of different vocalists. The gesturing in Karnatak contexts is akin to that of related North Indian styles, and indeed, the majority of research on gesture and Indian music has focused on North Indian practices, including Khyal and Dhrupad. Across these styles, gestures are not taught formally. Instead, the tendency to gesture in certain ways appears to be acquired implicitly during the lengthy learning process.[3,4] In Indian musical contexts, performer and audience gesturing has been analyzed to explore topics, including audience perception of metrical structure,[2] kinetic analogy through sound,[6] vocalization and gesture as parallel channels for melody,[3] metaphor, iconicity, and cross-domain mapping,[4] performance practice across cultural contexts,[7] and connections between physical effort and vocal sound.[5] Notwithstanding this body of work, questions remain regarding the nature of the coupling between gesture kinematics and vocal sound, namely, which aspects of each are structurally related to the other. Until recently, research on music-related gesture–vocal coupling has been hampered by a lack of appropriate statistical methods for assessing the coupling, but methodological progress has provided new solutions, which we employ here.

In this study, we ask what features of vocal sound and gesture kinematics are most closely coupled, and in what way. Second, we examine how this varies across performers and performances, and whether the nature of the coupling is affected by musical context, in particular across the melodic frameworks known as ragas (*rāgas*).[a] Such examination will contribute to understanding of what the gestures index or represent, and how, which has relevance for the question of why performers gesture as they do. Furthermore, the findings could have implications for broader debates regarding how gestures relate to vocal production through the physical connections between the two systems. These questions are all part of the general unsolved conundrum of why humans tend to move along with music (or play music along with movement) in complexly varied ways.[8–11]

The study has distinctly interdisciplinary foundations, and its theoretical background is drawn from two areas; the first from gesture studies, where there is an existing body of research addressing coupling between gesture and vocal production in speech contexts, the second from work within musicology and associated fields on relationships between music and movement, and also on the aesthetics of the Karnatak style.

[a] Ragas are melodic frameworks that include rules on which musical pitches can be played and in which order, the *gamakas* (ornaments) that should be played on those pitches, and the characteristic phrases that must be performed in order to properly express the raga. As a result, each raga is considered to have its own particular character or "color."

BACKGROUND IN GESTURE–SPEECH COUPLING

During speaking, the hands gesture not only to convey meaning through enacting, depicting, or symbolizing, but also by adopting a certain prosody (or melody) together with speech. When emphasis is given to a speech segment, the concomitant rising excursion in the fundamental frequency is often accompanied with a salient jerky movement of the gesturing hand.[12–15] This so-called beat-like quality of gesture—where manual movement synchronizes with sharp rises in pitch and other acoustic markers of emphasis—has traditionally been understood to be a cognitively acquired tendency.[15,16]

However, the upper limbs are attached to the torso, which is part of the respiratory-vocal system. It is known that simply moving your arms can interact with breathing cycles,[17] and sudden arm movements can change the intra-abdominal (and thus potentially subglottal) pressures due to recruiting a wider set of posture-stabilizing muscles during upper-limb movements.[18,19] Indeed, in nonhuman animals, such as flying bats and birds, vocalizations entrain to wing beats because the flying-related muscle tensions affect respiratory-vocal control.[20,21] Gesture–speech physics research has followed this line of thinking through for human manual gesturing, and found that chest kinematics are affected by upper limb movement, which in turn affect acoustic markers of emphasis during, for example, monosyllable utterances.[22] It has also been shown that standing (vs. sitting), moving higher mass effectors (two arms > one arm > hand), with higher de/accelerative movements leads to increased acoustic markers of emphasis of upper limb movement in vocalizing and fluent speech.[23,24] This supports the gesture–speech physics hypothesis that the beat-like quality of gesture should be understood as force-generating physical impulses, which recruit a wider set of (posture-stabilizing) muscles,[25,26] some of which that are implicated in respiratory-vocal control.[19,27,28] Thus, acceleration is an important kinematic marker of gesture–speech *kinetics*, as force transfers (i.e., physical impulses) are necessarily occurring when a body segment with a certain mass accelerates or decelerates over some time (note that force = mass × acceleration). Other kinematic variables, such as speed, or position, are less directly informative about the forces generated by gesturing. Thus, according to the gesture–speech physics thesis, acceleration is a key kinematic marker of force generation, and is predicted to be the most reliable parameter to understand coupling between gesture and the respiratory-vocal system. This has recently been supported by machine learning results showing that gesture acceleration is better predicted from speech acoustics as compared to gesture speed.[29]

The existing research suggests that there are physical tensegrity-related interconnections between the respiratory-vocal system and the upper limbs, which we predict are enacted in this real musical

context too. We study this here through an exploratory analysis based on acceleration being a marker for force transfers across the body that are primarily powered by muscle contractions (but also tensile equilibrating properties of connective tissues). If acceleration couples more strongly with acoustic variables than with speed or vertical velocity, for example, this would be consistent with an interpretation in which force transfer is a significant factor in the coupling of gesture and vocal production. However, we do not mean to imply that gesturing in this context is purely determined by physical coupling, only that it plays a *dispositional* role; it poises performers to move their arms in a particular way rather than another—but it does not physically obligate them in any way, and there can be individual differences in how performers cope with the physical constraints of gesturing and vocalizing at the same time. Therefore, we see such gesturing as also being a particular strategy that is "cognitively" acquired and may then also be further shaped in relation to other cross-modal perceptual mappings, specific cultural tendencies, and the aesthetic goals of the vocalist for a particular performance. This leads us to characterize the coupling observed as a neuro–bodily distributed aesthetic entanglement, which we will further explicate in the discussion. We suggest that it would be misguided to imagine a binary opposition between either aesthetic goals or biomechanical stabilities. Instead, we propose that performances should be viewed as arising out of multiple constraints brought into harmony by the performer.

BACKGROUND IN MUSIC-MOVEMENT CORRESPONDENCES AND KARNATAK MUSIC AESTHETICS

In seeking to account for observed connections between co-musicking movements and musical sound, some research has looked to the potential influence of perceptual cross-domain mappings found between movement and sound/music.[4,5,30] For example, studies using motion imagery paradigms have identified correspondences between change in fundamental frequency and change in vertical position; increase in loudness and movement toward the listener; and increase in loudness and the application of force.[31,32] Meanwhile, studies where participants are asked to move or trace in response to short music-like phrases have similarly found correlations between change in fundamental frequency and vertical position,[33,34] loudness and movement velocity,[35] and impulsive sounds and high acceleration peaks.[34] Through cross-modal interaction, performers' gestures can also influence audience members' perception of musical sound, biasing pitch perception,[36] and perceived duration.[37] Similarly, in speech contexts, seeing a beat gesture can change the perception of a co-occurrent syllable as if it were lexically stressed.[38]

Following such research, music is now increasingly understood as a multimodal phenomenon, wherein physical movement and musical sounds are intertwined in behavior and experience. Cross-modal correspondences are held to develop largely through repeated experience of regularities in the environment,[39] including accumulated experiences of physical interactions with objects and the sounds that result.[32,40] For example, we know that when we hit an object with more force, a louder sound arises. Such insights connect with ecological psychology perspectives on sound perception, in which it is proposed that people perceive the causes of sounds more immediately than their acoustic qualities.[41–43]

Vocalists thus perform in the context of such perceptual cross-modal correspondences and statistical regularities across the body and environment, evidence of which can be observed in their gesturing.[4,5,44,45] However, the likely impact of performers' aesthetic goals on gesturing also demands consideration. Unlike in experimental studies on cross-modality, for the current study, vocalists were not asked to trace or respond to music with movement, but rather they gestured spontaneously while focusing on performance goals related to the aesthetics of the Karnatak style, as they normally do when performing. Based on interviews with the vocalists conducted alongside the raga *ālāpana*[b] recordings made for this study, the goals of such performances are to express the character and beauty of the raga and thus move the audience. In Karnatak music, the character of each raga is conveyed through its characteristic phrases and motifs, which must be performed with specific patterns of emphasis and de-emphasis through modulation of loudness, pitch, duration, and timbre.[46–48] The correct and expressive performance of such phrases forms the basis for what is considered beautiful in the style—its aesthetic qualities.[46] Therefore, the required modulations of loudness, pitch, and duration within characteristic phrases and motifs should be conveyed to the audience. As music is a multimodal practice, wherein cross-modal interaction means that gestures can affect the perception of musical sound (as discussed above), both sound and gestural movement may contribute to the audience's experience of these qualities. Gestures can thus be viewed as contributing to the aesthetic experience of performances, and the influence of particular perceptual cross-modal correspondences on gesturing should be viewed in this context. Considering that vocalists have performance goals, it seems likely that they use such correspondences skillfully and unreflectively in performance, with a keen appreciation of what is effective in their aesthetic endeavor.

CURRENT STUDY

In this study, we examine coupling between gestures and vocal sound based on the following research questions and rationales.

Which couples most strongly with gesture kinematics: F0 or amplitude?

Following existing work on cross-modal perception discussed above, we expect to find cross-modal relationships between performers' gestures and vocal sound produced. But, which variables couple most strongly in this real musical context? The question is important because

[b] Raga *ālāpana* is a musical format without meter, where the performer extemporizes on a raga based on its existing characteristic phrases and raga grammar. In this format, nonlexical vocables are sung, rather than lyrics.

strength of coupling can provide insight into which sonic features are indexed by performers' gestures, and thus may also have an impact on audience members' perception of the music (considering the literature on cross-modal interaction discussed above)? Here, we examine the strength of coupling between kinematic variables and two acoustic features commonly implicated in studies on cross-modal mappings in musical contexts—change in F0 and amplitude.

What couples most strongly with vocal acoustics: acceleration, speed, or vertical velocity?

Following existing research on gesture–vocal coupling in speech and musical contexts, we expect acceleration to couple most strongly with acoustic variables. We aim here to provide a fine-grained analysis of coupling between kinematic and acoustic variables, looking not only at temporal coupling but also magnitude coupling. Such an analysis is important for what it can tell us about how performers index the sonic features—which kinematic features are most strongly implicated. In addition, a finding of acceleration being most reliably coupled with acoustics would be consistent with interpretations highlighting the salience of force, as this is more directly related to acceleration than the other kinematic features analyzed.

How do the couplings examined above vary across performers and performance types (different ragas)?

We ask whether individual performers have idiosyncratic modes of coupling between gesture and acoustics, or whether there are commonalities across performers? In addition, we seek to discover whether the raga performed has an effect on the quality of coupling. This question is stimulated by the fact that ragas are often considered by musicians to have particular characters or moods.[48]

As Karnatak vocal performance is a complex human behavior involving physical, cognitive, cultural, and aesthetic influences, we expect the results to be similarly complex, but we hope to identify some underlying trends in answer to our questions through our analysis of a large number of performances by four Karnatak vocalists who, as socially acknowledged expert performers, are taken to be indicative of current performance practice in the style.

The outline of this study is as follows. We first compare acceleration peaks with 3D speed and vertical velocity peaks, the latter two of which are kinematic variables with a high likelihood of entraining to musical features (based on research discussed above). We focus on studying temporal regions around peaks in movement as we know that gestures are intermittent in their activity, such that there are moments of vocalization without gesture that we should not average with moments of gesturing. Further, our analyses procedure is tailored for the study of time series that likely couple polyrhythmically due to the inherently different time scales that define each system (see Methods). We then perform a coupling analysis that focuses on the presence of temporal coupling, such that some acoustic fluctuation consistently occurs relative to the timing of a kinematic fluctuation. Then, we follow up with a magnitude coupling analysis, which quantifies the degree to which a gesture kinematic magnitude scales with the acoustic fluctuations. Finally, we analyze whether there is consistent variability in gesture–vocal coupling across ragas (melodic frameworks) or performers. For example, some performers may use one particular cross-modal mapping (e.g., vertical motion with F0 change) over another (e.g., speed and F0 change).

MATERIALS AND METHODS

Performances and performers

In total, 35 recorded performances of raga *ālāpana* were analyzed, covering eight different ragas,[c] lasting *M* (*SD*) duration = 327.54 (124.97) s, min-max duration = 99.47–620.37 s. It should be noted that raga *ālāpana* is performed without a metrical structure (musical meter) or steady beat, therefore, the question of entrainment to a regular, repeated beat does not arise in this musical format. Four right-handed vocalists participated in this study (two males and two females, *M* age = 36.5, *SD* age = 6.46). These vocalists, based in Chennai and Bengaluru, are all respected and currently active performers within the South Indian, Karnatak music community, each having a combined studying and performing experience of between 22 and 37 years.

Recordings

Audio recording

Sound was recorded at 48 Khz using Neumann KM184 condenser microphones.

Motion tracking

Motion tracking was performed with Xsens MVN Awinda (Xsens, the Netherlands; 60 Hz sampling); a full body inertial sensor motion capture system. We smoothed *x*, *y*, *z* traces with a zero-lag 30 Hz third-order Butterworth filter to reduce noise-related jitter. We extracted movement traces (*x*, *y*, *z* displacements) of the left and right wrists.

Video recording

Video recording was performed with a GoPro Hero4 camera at 50 fps.

Manual-vocal events measurement

Manual gesture events were annotated in ELAN.[49] The gesturing events were defined as a sequence of movement strokes and post-stroke holds in the gesture space in front of the performer. The start

[c] All four performers sung all eight ragas, apart from raga Bhairavi, which was not performed by one vocalist.

boundary of the gesture event was approximately determined as the moment when the hand finished its preparatory phase from rest position to gesture space. The end boundary of the gesture event was the moment when the gesturing hand retracted to the rest position. Thus, the gesture events did not include the preparatory and retraction phases from and to rest positions (a common approach in co-speech gesture coding).[50]

Acoustic measurements

For each acoustic measure x, we consider the absolute change ($|\Delta x|$) in magnitude (i.e., the absolutized first derivative of x with respect to time). The derivative is used as we are interested in whether kinematics couple with the dynamic changes in the acoustics (does acceleration couple with F0 movements). This is different from asking whether gesture kinematics tend to covary with high or low F0/amplitude.

Absolute change fundamental frequency (|ΔF0|)

The fundamental frequency is the main acoustic determinant of the perceived pitch of the sound. Fundamental frequency was extracted to a time series with a sample rate of 200 Hz using Praat.[51] Pitch tracks were hand checked for noise-related tracking problems (e.g., period doubling), and the ranges for each performance were adjusted in Praat accordingly. We smoothed F0 with an 8 Hz Hanning window and then computed the absolute change of F0 over time, henceforth, |ΔF0|. |ΔF0| is expressed in Hertz change per second.

Absolute change amplitude envelope (|ΔENV|)

The amplitude envelope (ENV) tracks gross intensity changes in the sound. Using a custom-written script (https://osf.io/6vjqn/), we extracted the amplitude envelope from the audio. To extract the amplitude envelope, we applied the Hilbert transform and took the complex modulus of the analytic signal, yielding a 1D time series.[52] We then smoothed the amplitude envelope using an 8 Hz Hanning window. This smoothing of the amplitude envelope should provide us gross information in the acoustics that couple at comparable time scales with that of kinematics, while ignoring very fine structured information in the amplitude signal. We downsampled the sampling rate of our ENV time series to 200 Hz. We rescaled the amplitude envelope within each performance from 0 to 1. We then computed the absolute change of amplitude envelope over time, henceforth, |ΔENV|. |ΔENV| is expressed in a rescaled amplitude envelope unit (or arbitrary units, a.u.) change per second.

Kinematic measurements

Velocity *z*

The negative or positive rate of vertical displacement (velocity in the z dimension) was obtained. Positive values indicate that the hand is moving up, and negative values indicate movement downward. This measure is especially of interest if there is an acoustic mapping onto the vertical dimension, for example, positive change in F0 (increase in Hz) is reflected with an upward gesture. We express velocity z as a vector[d] quantity in centimeters per second.

Speed

3D speed of a particular body segment was calculated from individual x, y, z velocity (v) components, $s = \sqrt{vx^2 + vy^2 + vz^2}$, as provided by the motion tracking system. When speed is higher, it reflects that the body segment is moving faster in an arbitrary direction, and when speed is zero, there is no movement. Speed cannot be negatively valued. Therefore, we express speed as a scalar quantity in centimeters per second.

Acceleration

Acceleration was derived by quantifying the change in 3D speed over time (i.e., the first time derivative of s). Thus, in contrast to speed, acceleration can be negative and positive, and zero acceleration reflects that speed is constant. Hence, we express acceleration as a vector quantity in centimeters per second squared, indicating that over each second, there is a negative (deceleration) or positive (acceleration) change in speed.

Measurement aggregation

Acoustic measures and kinematic measures were aggregated using a custom-written processing script (https://osf.io/q3rxa/). We upsampled the motion tracking data using linear interpolation from 60 to 200 Hz, which was then aligned in time with the acoustic measures (already sampled at 200 Hz). All our coupling analyses take this acoustic + kinematic time series as their input.

Gesture–vocal temporal and magnitude coupling analysis

All analyses were performed in R (version 4.0.3), the script can be found here (https://osf.io/3mquh/). We investigate gesture–vocal coupling by first assessing how well vocal trajectories can be modeled around 1 s of a kinematic peak, and how vocal trajectories differ depending on whether the kinematic peaks are of higher or lower magnitude. We use generalized additive modeling (GAM), with `R`-package `gam`,[53] to test for nonlinearities over time in the form of consistent peaks in acoustics around kinematic peaks, which allows for identifying temporal coupling at potential lags (Figure 1). Significant time lags are, for

[d] Meaning that negative (downward) or positive (upward) velocity is informative for the direction of movement rate of displacement in the z-dimension.

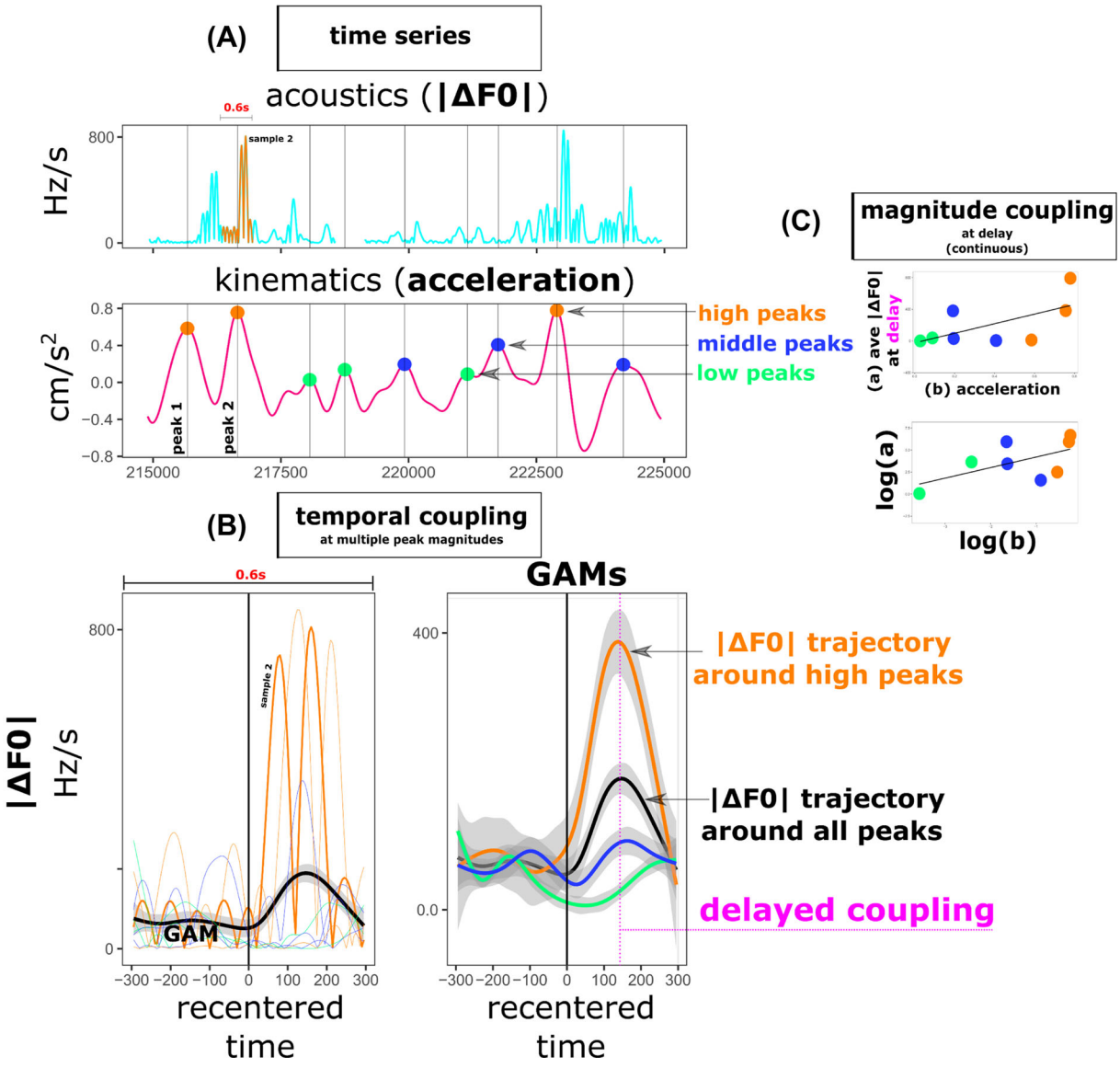

FIGURE 1 The upper panels (A) show the time series from a gesture–vocal event by performer 3 (see https://osf.io/shfbc/). The uppermost panel includes the absolute change in F0 (|ΔF0|), where higher values indicate greater changes (negative or positive) in the vocalization's F0. The lower panel of A shows the concomitant gesture acceleration time series, where the positive peaks are identified and given a magnitude category based on 33% quartiles (three lower peaks, three middle peaks, and three high peaks). The time series are then analyzed for temporal coupling (while taking into account magnitude peaks) as shown in B. For the temporal coupling analysis, we prepare the time series for GAM by sampling vocal time series around an interval (here 600 ms) centering the kinematic peaks. An example of such a sample is given in B, which shows peak 2 of the time series in panel A (annotated as "sample 2"). If we repeat this process for all peaks, we can generalize over the vocal trajectories while taking into account magnitude, and this results in GAM fitted (nonlinear) slopes as given in the right panel of B. It can be seen that there is a consistent pattern for this single gesture event such that after about 130 ms, a positive peak in the acceleration |ΔF0| follows, that is, there is a delayed temporal coupling. It can also be seen that this general pattern (given in black) is much more pronounced for the high magnitude peaks, followed by the middle peak, and then the low peak, which suggests a role for magnitude of the peak in establishing temporal coupling. (C) To further assess the magnitude coupling, we establish from the GAM the delayed coupling in milliseconds, and then take average samples of |ΔF0| and relate this continuously to acceleration peak magnitude using linear mixed regressions.

example, expected if there is some kind of dynamic neurophysiological feedback between gesture and vocalization trajectories, or when there is a biomechanical coupling between gesture and vocalization. For example, anticipatory postural muscle adjustments that are implicated in respiratory control often occur at about 70 ms before and after the upper-limb-induced postural perturbation,[27,54] and in previous work, we have found in complex full-sentenced speech that peaks in F0 and amplitude envelope are found 100 ms before the maximum extension of the lower arm around the elbow joint.[23] Similar reflexive delayed loops could involve sudden increases in subglottal pressure due to external mechanical loading on the chest leading to reflexive laryngeal counter adjustments after about 30 ms.[55]

We further added fixed effects of different kinematic peak magnitudes to the GAM model so as to provide an initial test for magnitude coupling (does acoustic output generally scale with kinematic peak magnitude?) and visualize how kinematic magnitude affects the temporal coupling of kinematics with the vocal trajectory (does temporal coupling especially arise when the kinematic peak reaches a certain magnitude?). We refer to Figure 1 as a graphical explanation of our temporal analysis approach.

Specifically, for each gesture event, we identified all positive peaks (in the case of scalar quantity, such as speed) or negative *and* positive peaks (in the case of vector quantities, such as acceleration and vertical velocity) in the time series. Kinematic peaks were determined using a peak-finding algorithm implemented by R-package `pracma`.[56] In order to capture sufficient variability in the magnitude of acoustic peaks to be related to the kinematic peak, the peak-finding algorithm was not thresholded (e.g., minimum magnitude of the peak), though positive or negative peaks needed to exceed the 0 boundary. Thus, we also have peaks that are of relatively minor magnitude, next to more pronounced kinematic peaks (Figure 1). To make a distinction between different magnitudes of the peaks and their relation to temporal coupling, we initially distinguish between low, middle, and high magnitude peaks, which were determined for each performer separately by identifying the lowest (0–33% quantile), middle (33–66%), and highest (66–100% quantile) peaks for that performer. In a follow-up analysis, we quantify magnitude coupling continuously, and these arbitrary cutoffs are merely used for assessing context-dependent effects of magnitude coupling with temporal coupling.

For modeling, we performed GAM to assess whether the kinematic measures reliably coupled with the acoustic measures while factoring out the variance attributed to performer and performances (i.e., different ragas). GAM is a type of nonlinear mixed regression,[57] which uses a set of basis functions to optimally model a particular nonlinear relation between variables. To reduce chances of overfitting, the GAM algorithm penalizes more complex nonlinear functions relative to variance explained.

After the GAM analysis focused on modeling the temporal structure of vocal trajectories, we performed a mixed linear regression analysis specifically tailored to the precise quantification of magnitude coupling, where we assessed the continuous kinematic magnitude relative to the acoustic change (e.g., how much change in vocal acoustics should we expect per cm/s increase in speed?). As it is possible that magnitude coupling is not simply linear, we also assess nonlinear relationships by regressing variables after a log-log transformation (see Results for further details).

Comparisons to other types of analyses

Note that our analysis procedure differs in some ways from other common approaches to assessing synchronization processes in music and movement.[34,58,59,60] Cross-wavelet analysis is currently a popular approach for analyzing (multiscale) synchronization between two time series of a similar nature, such as movement produced by two persons.[58,60] However, changes in vocal acoustics are of several magnitudes faster in oscillation period than the relatively slower manual gesture system (e.g., compare |ΔF0| and the acceleration time series in Figure 1). Hence, the two systems operate on inherently different time scales and are likely to couple polyrhythmically,[61] or indeed in pulses,[22] and it is then less ideal to perform an analysis that is designed to assess continuous coupling within time scales between (multiple) oscillations that have a comparable period. Further, our analysis is sensitive to lagged synchronization processes, similar to cross-correlation analysis or cross-recurrence quantification analysis.[62] However, the current GAM approach (combined with linear mixed regression) provides a statistical tool for disentangling random effects of performer and performance, from nonlinear main effects of time relative to peak kinematics and fixed effects of kinematic magnitude.[57] A drawback of the current approach is that it is not possible to infer which system is likely causing (in a statistical sense) some effect in the other, and approaches, such as Granger causality analyses, are an interesting further avenue of inquiry for the current research.[62]

Performer and raga difference in temporal and magnitude coupling analysis

Our main temporal and magnitude coupling analysis is focused on whether we can generalize over performers and performances the way that gesture couples with vocalization. Of course, this might obscure interesting differences between performers or what is being performed (i.e., which raga). We will, therefore, further explore performer and raga-dependent differences by summarizing all GAM model fits of the different gesture–vocal coupling using dimensionality reduction (principal component analysis; PCA).

RESULTS

Descriptive performances

Gesture–vocal events

There were an *N* total of 1630 gesture events detected by the annotator. Performer 1 had a mean gesture event rate per second of $M = 0.156$ ($SD = 0.012$), performer 2 $M = 0.170$ ($SD = 0.018$), performer 3 $M = 0.145$ ($SD = 0.019$), and performer 4 $M = 0.167$ ($SD = 0.031$). Gesture–vocal events had a mean duration of $M = 4.67$ s ($SD = 2.75$); performer 1 $M = 5.08$ s ($SD = 2.51$), performer 2 $M = 4.43$ s ($SD = 2.41$), performer 3 $M = 4.50$ s (SD = 2.68), and performer 4 $M = 4.88$ s ($SD = 3.13$).

Gesture kinematics and vocal acoustics

In Table S1, we report descriptive information about gesture kinematics (e.g., average peak velocity of a gesture). Table S2 provides

TABLE 1 Generalized additive modeling coefficients for |ΔF0|

Models	Parametric effectsPeak magnitude	Parametric effects *p*-value	Smooth components	*F* [edf, ref.df]	*p*-value	Deviance explained
|ΔF0| ~ velocity *z* (positive peaks)	Low versus high: −122.28 Middle versus high: −62.22	<0.001 <0.001	Recentered time: low Recentered time: middle Recentered time: high random (Perform, Raga)	3 9.1 [7.92, 8.99] 106.6 [8.43. 8.91] 456.0 [7.92, 8.70] 3221.6 [27.99, 28.00]	<0.001 <0.001 <0.001 <0.001	12.00%
|ΔF0| ~ velocity *z* (negative peaks)	Low versus high: −98.98] Middle versus high: −46.39	<0.001 <0.001	Recentered time: low Recentered time: middle Recentered time: high random (Perform, Raga)	15.0 [6.81, 7.92] 33.6 [7.51, 8.45] 214.3 [8.85, 8.99] 2657.4 [27.99, 28.00]	<0.001 <0.001 <0.001 <0.001	9.63%
|ΔF0| ~ speed (positive peaks)	Low versus high: −101.34 Middle versus high: −42.14	<0.001 <0.001	Recentered time: low Recentered time: middle Recentered time: high random (Perform, Raga)	325.5 [6.78, 7.91] 432.4 [8.78, 8.99] 535.8 [8.64, 8.96] 3180.1 [27.99, 28.00]	<0.001 <0.001 <0.001 <0.001	12.12%
|ΔF0| ~ acceleration (positive peaks)	Low versus high: −166.63 Middle versus high: −45.16	<0.001 <0.001	Recentered time: low Recentered time: middle Recentered time: high random (Perform, Raga)	166.5 [7.23, 8.26] 1007.8 [8.32, 8.87] 2537.1 [8.49, 8.93] 3365.3 [27.99, 28]	<0.001 <0.001 <0.001 <0.001	12.9%
|ΔF0| ~ acceleration (negative peaks)	Low versus high: −90.76 Middle versus high: −27.57	<0.001 <0.001	Recentered time: low Recentered time: middle Recentered time: high random (Perform, Raga)	1960.7 [9.41, 8.90] 1084.6 [8.28, 8.86] 1960.7 [8.41, 8.90] 3554.1 [27.99, 28.00]	<0.001 <0.001 <0.001 <0.001	12.20%

information on performers' vocal acoustics for (change in) fundamental frequency and amplitude envelope for vocal events with and without gesticulation. It can be seen from Table S2 that across the board, there is higher change in vocal acoustics (|ΔF0| and |ΔENV|) during gesturing versus no gesture, as well as higher amplitude envelope, but not higher fundamental frequency.

Temporal coupling analysis

Table 1 provides the GAM modeling coefficients with each model's explained deviance for |ΔF0|. Please see Table S3 for the GAM results for amplitude envelope, which showed generally poorer modeling performances (<6% deviance explained), though slightly better fit for vertical velocity (6.03%) as compared to speed (4.38%) or acceleration (<5.43%). Figure 2 provides the fitted trajectories for all models. As a sanity check, we also model the kinematic trajectories, as separated by magnitude of the kinematic peaks (low, middle, vs. high magnitude). The related vocal trajectories show differences per magnitude of the peak, with (1) higher baselines for higher magnitude kinematic peaks (indicating a generally higher |ΔF0| during the 1 s around the kinematic peak) and more pronounced peaks (indicating a heightened |ΔF0| at a particular moment relative to the peak in kinematics; i.e., temporal coupling is more pronounced for higher magnitude of the kinematic peak). From Figure 2, it can be seen that especially for lower speeds, vocal trajectory peaks are aligned with no delay, and seem to become a little bit more delayed when coupled for higher speeds (as indicated by a shift of the peaks of the trajectories, such that surges in vocal changes occur before peaks in speed). For acceleration peaks, there is a delayed relation such that positive peaks in acceleration tend to be followed by a sharp increase in |ΔF0| of about 200 ms. Note further that for the deceleration peaks (negative peaks in acceleration), the |ΔF0| occurs before such peaks with about −225 ms. Graph S2 (https://osf.io/szy8t/) shows the GAM fitted trajectories for |ΔENV|.

We noticed from inspecting the GAM results reported in Table 1 that the model fits diverge especially for higher magnitude peaks (deviance explained low < medium < large), as indicated by *F*-values for each fitted trajectory per peak magnitude. As the *F*-statistics are not directly comparable between models (e.g., due to different degrees of freedom), we excluded the low and middle magnitude kinematic peaks and reperformed the GAM analysis to discern whether temporal coupling differences are especially pronounced for higher kinematic magnitudes. When focusing only on the high kinematic peaks, acceleration (positive peaks [12.7%] and negative peaks [12.7%]) more clearly outperforms velocity *z* (positive peaks [11.9%] and negative peaks [8.65%]) and speed (8.28%) in explaining deviance in |ΔF0|. Note these models are all generalizations over performers, and there might be considerable individual differences underlying these patterns. In the final "Individual and performance differences" section below, we report on performer differences in temporal coupling.

Quantifying magnitude coupling

We found that sudden vocal changes occur around peaks in speed, acceleration, and deceleration, especially for |ΔF0|, and secondarily |ΔENV|. We thereby show clear temporal coupling with kinematics. We also found an indication that such peaks in vocal changes scale with the magnitude of kinematic peaks, providing clear evidence that

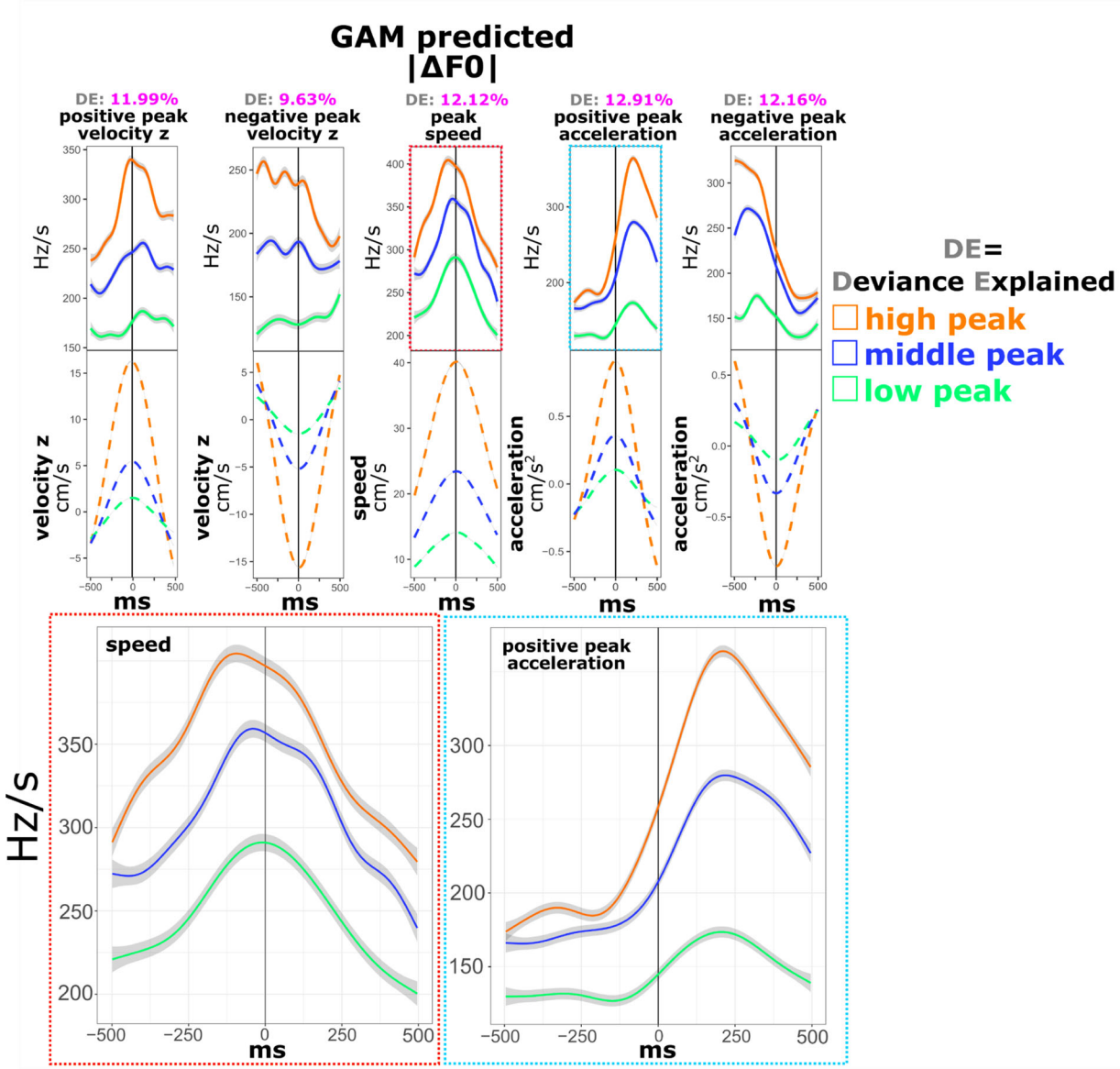

FIGURE 2 Generalized additive modeling around peak kinematics. The upper row shows the GAM predicted vocal trajectories around a kinematic peak (and reported in Table 1). The row below shows the kinematic trajectory relative to a positive or negative peak and provides a sanity check that we have normalized time correctly such that at time 0, there is a kinematic peak of a particular magnitude. The lowest left panel shows the vocal trajectories for speed and the lowest right panel shows enlarged vocal trajectories for positive acceleration.

magnitude coupling is also occurring. However, we should model the magnitude of kinematic peaks continuously relative to the magnitude of vocal changes, to provide a strong estimate of the differences in magnitude coupling between kinematic features. We do this for |ΔF0| as this was the acoustic feature most strongly coupled to kinematics.

For this analysis, we use information from the GAM models to sample acoustic magnitude around kinematic peaks. For speed and vertical velocity, the acoustic peaks are occurring closely around the moment of the kinematic peak, so we average local maxima occurring in an acoustic sample for each kinematic peak event at −25 to +25 ms. For positive peaks in acceleration, local maxima in acoustics are obtained and averaged at +200 ms, so we average local max values around +175 to 225. For negative peaks in acceleration, local peaks in acoustics are obtained at −250 ms, so we take the average max values around −225 to −275.

In Figure 3, each scatter point represents a magnitude of a peak in kinematics (*x*-axis) versus the mean of the 50 ms acoustic sample of |ΔF0|. When negative kinematic peaks are concerned, we absolutized them (e.g., |velocity *z*|, |acceleration|), such that higher values (rather than more negative values) indicate higher magnitude peaks. The lower panel shows the raw values and the upper panel shows a log-log plot. There were signs of strong nonlinearity and non-normality in the untransformed data. However, as can be seen, after the log-log transformation, we obtain a clear linear relationship between the variables, indicating a power-law function, which is a common scaling relationship in biomechanics. In the lower panel, we show

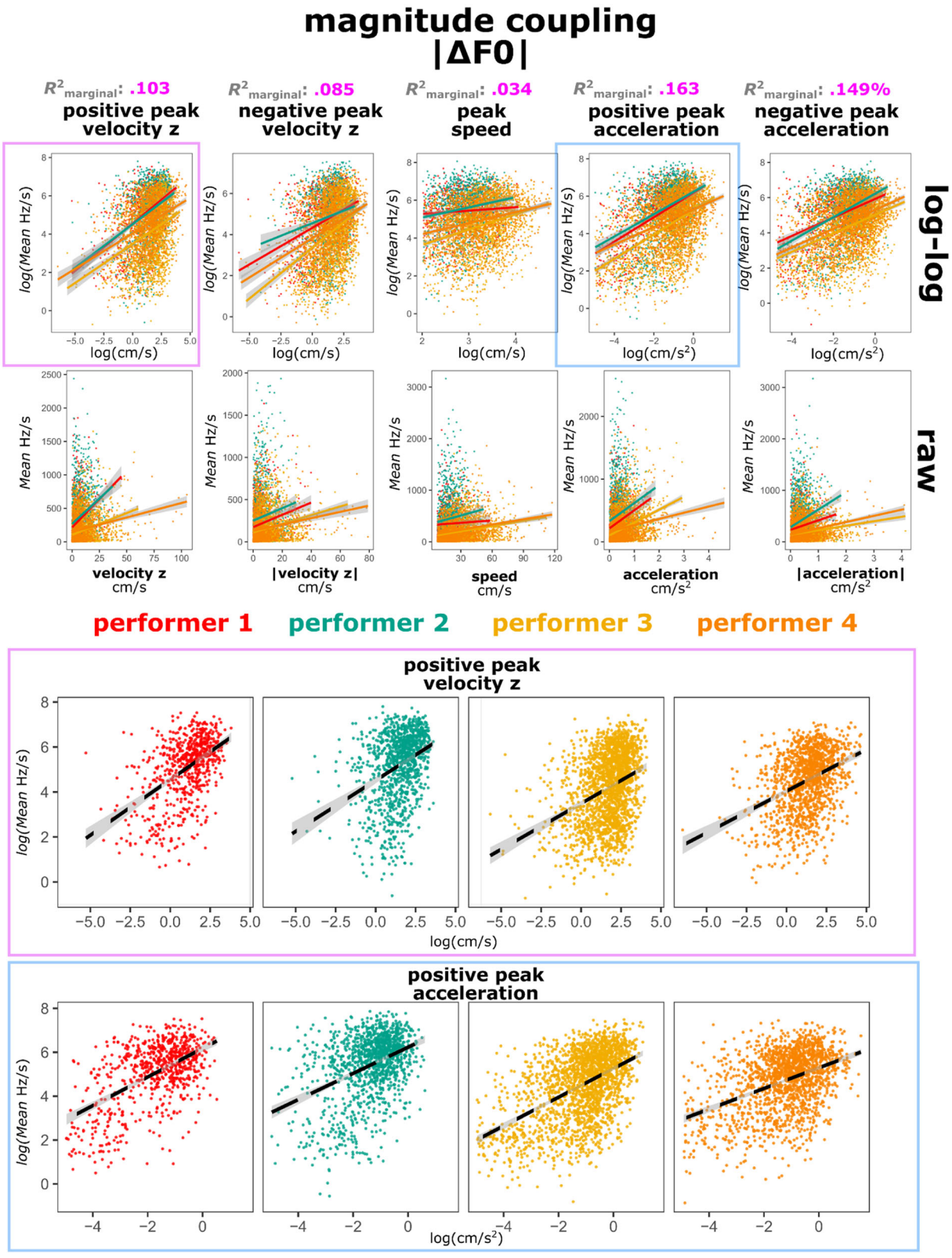

FIGURE 3 Scatterplots for magnitude peaks in kinematics versus average |ΔF0| for log-log scaled values (first row of plots) and raw values (second row of plots). The lower panels are an enlargement of plots in the first row (for positive vertical velocity and positive acceleration), separated out for performer.

TABLE 2 Linear mixed regressions for assessing magnitude coupling

Models DV: log(|ΔF0|)	*b* *t* (df)	*t* (df)	*p*-value	Fitted power law $y = \alpha * \text{peak}^b$	*Marginal* R^2
Intercept	−0.26	−15.88 (5686)	<0.001	$\|\Delta F0\| = 0.77*\text{peak}^{0.34}$	0.103
Log Peak velocity *z* (positive)	0.34	14.60 (5686)	<0.001		
Intercept	−0.04	−0.12 (5296)	0.902	$\|\Delta F0\| = 0.96*\text{peak}^{0.29}$	0.085
Log Peak velocity *z* (negative)	0.29	5.78 (5296)	<0.001		
Intercept	2.66	44.28 (4923)	<0.001	$\|\Delta F0\| = 14.29*\text{peak}^{0.08}$	0.033
Log Peak speed (positive)	0.08	4.49 (4923)	<0.001		
Intercept	−3.16	−12.80 (5733)	<0.001	$\|\Delta F0\| = 0.04*\text{peak}^{0.37}$	0.163
Log Peak acceleration (positive)	0.37	10.23 (5733)	<0.001		
Intercept	−2.94	−15.88 (5686)	<0.001	$\|\Delta F0\| = 0.05*\text{peak}^{0.32}$	0.149
Log Peak acceleration (negative)	0.32	14.596 (5686)	<0.001		

for each performer separately the relation between acceleration and acoustics.

Table 2 shows the main results of the linear mixed effects modeling. Mixed linear regression results are given for models with random intercept and slope for raga nested in the performer. The dependent variable is the logarithmically transformed average acoustic sample of |ΔF0|, and the independent variable is the log-transformed kinematic peak. When performing the models with a log-log transformation, the intercept *I* is informative about the multiplicative constant of the power law (where $\alpha = e^I$). The beta coefficient is informative about the power relationship, such that a $b \sim 0.33$ indicates that for every doubling (i.e., 100% increase) of the kinematic peak magnitude, there is a 33% increase in |ΔF0|.

We consistently find that |ΔF0| is best predicted by all kinematic variables through log-log transformation as is also evident from Figure 3, indicating a nonlinear scaling between sound and kinematic magnitude. Log |ΔF0| is best predicted by the log acceleration as indicated by the higher marginal effect sizes (>14.9%) relative to other kinematic variables (<10.3%) that quantify the fixed effects contribution in explaining the data (Table 2). From these marginal effect sizes, we conclude that gesture acceleration (as compared to speed and vertical velocity) is the best predictor for the magnitude of |ΔF0|.

Note, that the *conditional* effect sizes are also informative about the by-performer and by-raga modulation of the kinematic effects. In other words, how much variance is explained by the model if we unfix the fixed effects coefficients so that they vary per raga and performer, for example, allowing the model to parameterize a more or less extreme kinematic effect depending on raga and performer. Interestingly, although we found that positive peaks in vertical dimension had a low fixed effect size (10.3%), suggesting a poor generalizability of the effect, the conditional effect size is quite high (46.56%), making it clear that coupling between vertical velocity and sound is very much dependent on modeling such effects variably depending on raga and performer. The conditional effect sizes for negative vertical velocity (40.20%), speed (35.55%), and acceleration (positive peaks: 43.8%, negative peaks: 35.45%) were lower than that of positive vertical velocity. We interpret this as performers using positive vertical motion in a between-subject variable way, but reliably so across the performances within each performer.

We conclude from all this that the gesture–vocal coupling is generally best described by acceleration as is evident from the higher fixed effect sizes. But, we also observe that depending on performer and raga, there can be high predictive performance by positive vertical velocity. In other words, acceleration-|ΔF0| is evident across ragas and performers, but vertical mapping is more variably an important mode of gesture–vocal coupling (conditional on a particular raga or performer).

Individual and performance differences

Gross generalities between performers' gesture–vocal coupling can obscure larger individual differences in gesture–vocal coupling styles. The previous magnitude coupling analysis already provided some information about differences between performers and ragas (as particularly evident in the conditional effect sizes for vertical velocity), and in this final Results section, we further visualize and analyze the cross-performance and cross-performer variability that underlies our general results.

As an indication of the individual differences of performers and ragas in the temporal coupling analysis, we have refitted GAM trajectories for |ΔF0| by performer and raga for vertical velocity *z* and acceleration separately (Figure 4).

To investigate this variability as seen in Figure 4 (and Figure 3), we extract information from our models for each gesture–speech coupling relation using GAM (temporal coupling) and linear mixed effects model (magnitude coupling) separately for each participant, so as to determine whether there are individual differences across performances in terms of which kinematic variable couples most with vocal acoustics (for simplification, we reduce our analysis to only |ΔF0|). To visualize this variability across many models, we use a dimensionality reduction PCA to plot potential differences in performances and/or

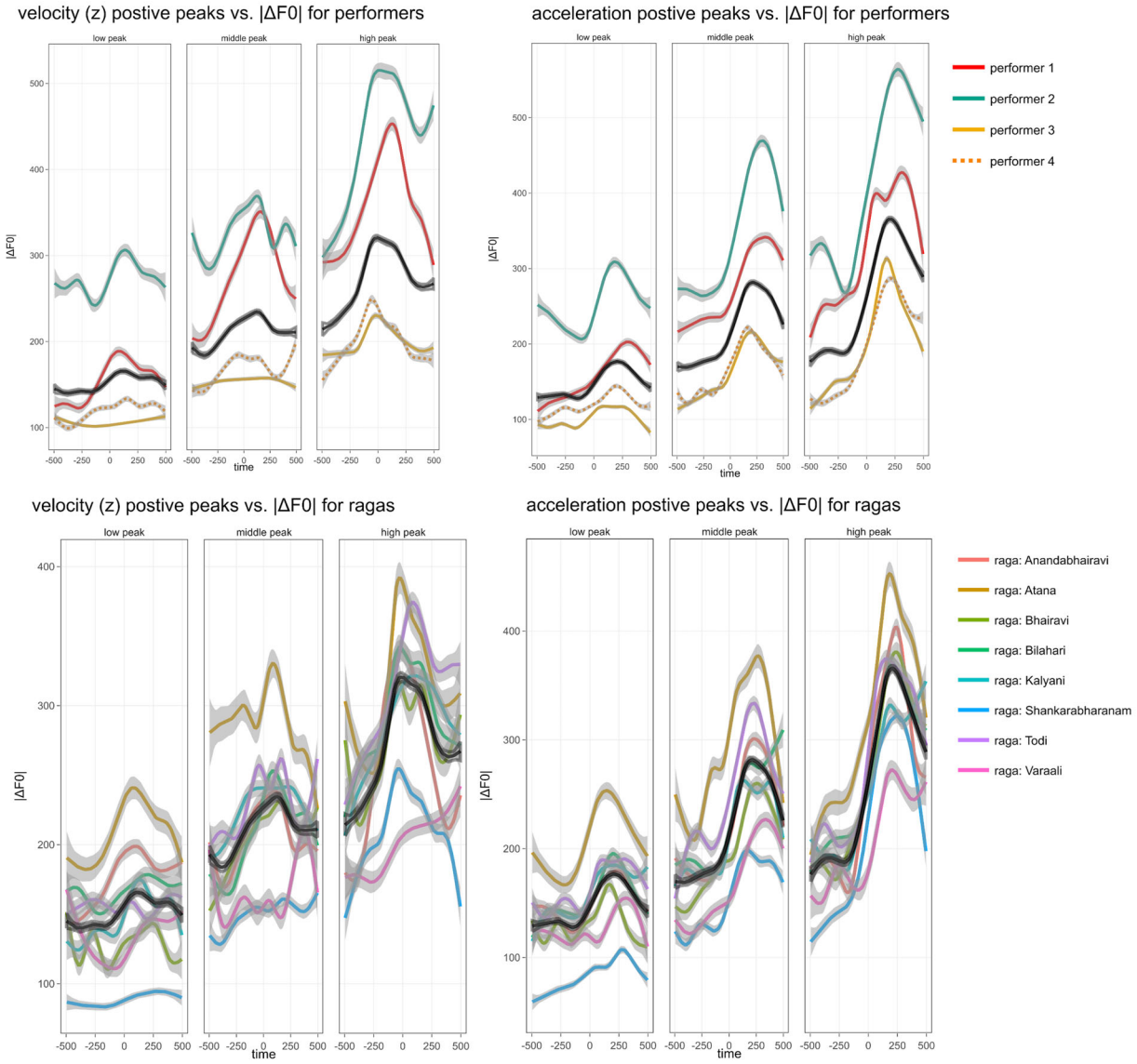

FIGURE 4 |ΔF0| GAM trajectories are plotted per performer (1 through 4) and for the different ragas. The black colored lines reflect the general trajectory over all the performers and ragas. Gray shaded areas indicate the standard error. Please note that interactive versions of the current graphs for velocity *z* (positive peaks; by performer and raga) and acceleration (positive; by performer and raga), as well for velocity *z* (negative; by performer and raga), speed (by performer and raga), and acceleration (negative; by performer and raga) are provided on our supplemental page on the Open Science Framework.

performers (Figure 5). Figure 5 left plot shows that for the raga *Shankarabharanam (Śaṅkarābharaṇam)*, there is relatively lower kinematic–vocal coupling, especially for acceleration, deceleration, and speed coupling. Further, raga Todi (*Toḍi*) seems to have a high variability of gesture–vocal coupling. The right plot shows the same values but colored by-performer. Performer 1 seems the most likely to exploit the vertical dimension in gesture–vocal coupling, while performer 2 is more likely to couple with acceleration and speed in a very consistent way (lower variability). By comparing Figure 5 plots 1 and 2 in terms of the degree to which the raga or performer occupies similar regions of the plot, it seems that performers are more consistent in their gesture–vocal coupling, compared to within ragas, where there is less consistency.

Table 3 provides an example of difference, comparing velocity *z* positive peaks and acceleration positive peaks per performer. For the temporal coupling, the explained variance is based on the GAM's deviance explained of |ΔF0| per kinematic variable. For the magnitude coupling, the explained variance is based on the marginal R^2 of the linear regression given for the sample of |ΔF0| per kinematic variable (also expressed as percentage of variance explained).

There is considerable variability in gesture–vocal coupling across ragas and performers. To provide a quantitative indication of how well the variability is captured by performer or raga category, we set up a machine classification task using R-package `caret` (Kuhn[63]). For three different seed initializations, we trained a random forest classifier on randomly selected 50% of the data, so as to then predict on the other

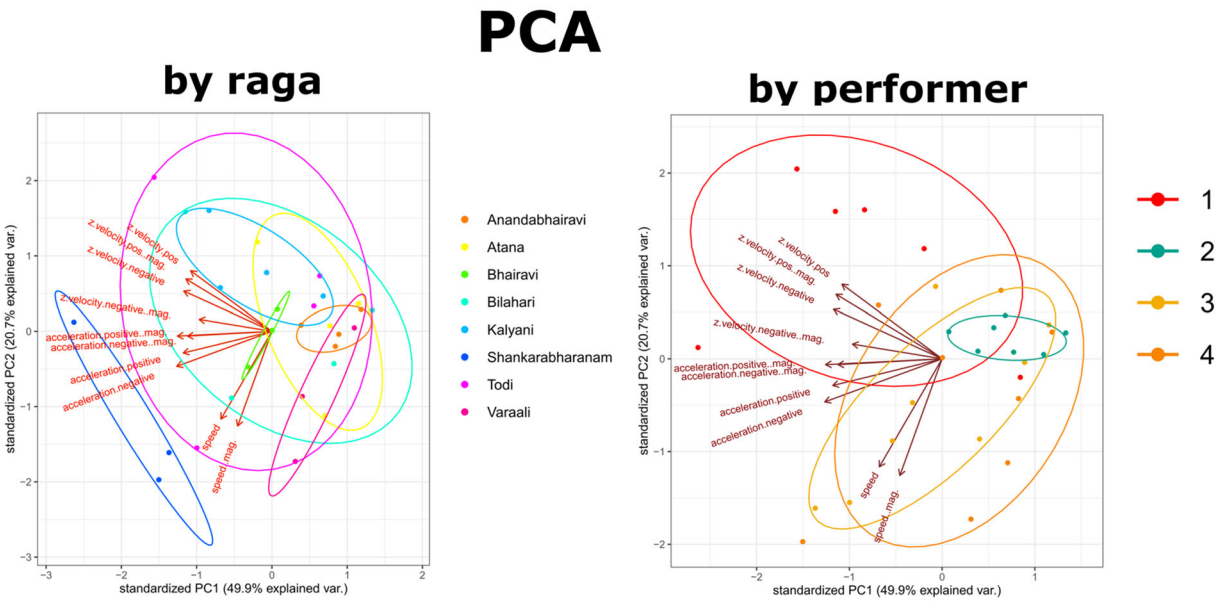

FIGURE 5 PCA biplots for raga (left plot) and performer (right plot) are shown, indicating how much of the variability in the gesture–vocal coupling (as indexed by deviance explained by GAM model for each performance) is structured according to raga or performer. The arrows indicate the dimension of variability. For example, when there is a positive value in the direction of z.velocity.pos, this indicates that there was relatively more deviance explained by a GAM model for |ΔF0| for that performance. Note also that some arrows are aligned, meaning that those dimensions are correlated (e.g., the arrows of acceleration.positive and acceleration.negative are aligned, indicating that the deviances explained are correlated).

50% of the data the performer or raga. We scaled and centered the data before training, and used a repeated cross-validation method for training (for code, see: https://osf.io/3mquh/). Though our results should be carefully interpreted given that we do not have many datapoints, we obtained that the classifier could not differentiate ragas on the basis of the model results, yielding an average classification accuracy of 17.95% (Table S4). This suggests that gesture–vocal coupling is not sufficiently different across ragas. We did, however, find that a machine learning (ML) classifier could more reliably predict the performer class in the testing set (52.38% accuracy), suggesting that performers were somewhat stable in their gesture–vocal coupling (Table S5).

DISCUSSION

In this article we studied coupling between gesture kinematics and vocal acoustics in a South Indian style of musical performance that is characterized by multimodal expression. We asked which kinematic variables that are in turn informative about physical effort and impulse (acceleration), amount of movement per unit time (speed), and vertical mapping (vertical velocity), are most coupled with changes in vocalization acoustics (fundamental frequency and amplitude envelope). Across the board, kinematics were more strongly temporally coupled with F0 rather than amplitude. This is interesting as although amplitude variation is important in music performance, F0 is arguably a more perceptually significant acoustic variable for characterizing the nature and structure of a melody. The gestural indexing of musical pitch may have aesthetic and communicative benefits in this style. This is particularly the case, considering the aesthetic importance in Karnatak music of conveying subtle nuances of pitch movement, along with the quality and rhythms of this movement, including which pitches are emphasized or de-emphasized (as discussed above in the background on music section).

We find that acceleration has the highest predictive performance for modeling the timing and magnitude of nearby change in F0, both for temporal coupling (are acoustic fluctuations timed relative to kinematic peaks?) and magnitude coupling (is magnitude of the acoustic fluctuations scaled to the magnitude of the kinematic peak?). However, for certain performers, the coupling of upward movements with changes in F0 was also evident. To understand this individual and by-raga variation, we also followed up with a machine classification analysis, where we tried to predict the performer or raga based on gesture–vocal coupling modeling results across kinematic and acoustic variable relations. We obtained that the variability in gesture–vocal coupling is more determined by performer than by a particular raga, though in general, the classification accuracy was poor, perhaps due to lack of data, and perhaps also suggesting that gesture–vocal coupling is something that is more richly and creatively varied by the performer.

The nature of gesture–vocal coupling

Gesture–vocal coupling strength

In general, kinematics and acoustic relations were weakly associated in time (explaining less than 13% of the variance) to moderately

TABLE 3 Individual differences in deviance explained GAM (temporal coupling) and marginal R squared for linear model (magnitude coupling)

	Performer 1 Temp. coupling Mag. coupling	Performer 2 Temp. coupling Mag. coupling	Performer 3 Temp. coupling Mag. coupling	Performer 4 Temp. coupling Mag. coupling	Mean (SD) Temp. coupling Mag. coupling
Velocity *z* (positive peaks)	11.56%	4.98%	4.44%	4.48%	6.05% (3.38%)
	25.11%	10.11%	11.83%	12.87%	14.44% (9.33%)
Velocity *z* (negative peaks)	8.26%	3.37%	5.29%	3.72%	5.01% (2.89%)
	14.32%	3.38%	14.55%	13.03%	11.38% (7.16%)
Speed	5.61%	4.46%	6.41%	6.03%	5.67% (1.64%)
	1.67%	3.67%	9.02%	9.22%	6.27% (6.09%)
Acceleration (positive peaks)	9.13%	6.91%	8.13%	6.59%	7.61% (2.25%)
	28.50%	14.13%	24.53%	17.26%	20.83% (10.05%)
Acceleration (negative peaks)	7.78%	5.56%	6.69%	6.66%	6.67% (2.21%)
	18.71%	16.72%	16.13%	18.82%	17.55% (6.95%)

associated in magnitude (explaining less than 16% of the variance), with increases in model performance when these effects were allowed to vary across performer and raga (up to 46%). However, based on the fixed effect sizes, we should be careful in concluding that gesture and vocalization are weakly associated. Namely, these percentages need to be weighted against the obvious fact that vocal acoustics is primarily structured by the melody's internal syntax and dynamic patterning, which need not be shared by gesture kinematics, and sometimes, simply cannot be shared by gesture because of operating on another time scale. For example, very fast vocal inflections might not couple with gesture kinematics due to the sheer inertia of the upper limbs, which make them too slow to reflect the acoustic structure in movement.[64] Thus, we argue that there is a theoretical (and unknown) upper limit for gestures to scale with acoustic fluctuations due to the different nature of the systems. The current effect sizes would ideally be expressed in relation to that theoretical upper limit. We should also consider that gesture–vocal coupling might have some *aesthetic* upper limit. Namely, it seems unlikely that performers aim for their gestures to act as a perfect linear transformation of their vocal patterning. Rather, gestures seem to be brought forth because they have their own unique dynamics and qualities. Indeed, in some cases, gesture is likely to be recruited precisely because it has features or affordances that are *not* shared with the vocal system. For example, when holding a long static note, one can occasionally see Karnatak vocalists continue to gesture the tracing of a straight horizontal line or static holding position while taking a breath, producing the impression of a (physical) continuation when in fact the voice has paused momentarily.

Gesture–vocal coupling scaling relations

Our magnitude coupling analysis showed that gesture–vocal coupling follows a power relationship. Namely, for every 100% increase in acceleration peak (negative or positive), about 32–37% increase in fundamental frequency change was found at a lag of 200 ms. This nonlinear gesture–vocal relationship was a positive power law relationship in the region of 1/3. Note, nonlinearities in general,[65] and power law scaling relationships specifically,[66] are commonly observed in human movement, for example, in the relation between curvature and tangential velocity of manual and articulatory movements, where a negative ⅓ power law relationship is typically observed.[66,67] Interestingly, as Levin notes,[26] scaling relations between a force perturbation on the body and its effects on more peripheral elements of the tensioned bodily system are also likely to be nonlinear, such that a more extreme perturbation (e.g., 2× higher gesture acceleration peak) can be attenuated by the bioarchitecture as its tensioned setup allows the distribution of forces over multiple elements of the system (leading to only a one-third acoustic effect). Though this finding aligns with those in human movement and biomechanics studies, the current magnitude coupling should not be treated as a power law in a strict 1:1 sense, as the acoustic fluctuations in our findings more variably scaled to kinematics, as compared to the classic negative one-third power law observed between movement curvature and speed, which almost pattern as a single-valued function.[66] Nevertheless, it is interesting that we obtain this novel multimodal power relationship between a kinematic and an acoustic variable, which should serve as a promising basis for further research that could connect principles from human movement science with research on the human voice.

Gesture–vocal coupling from a gesture–speech physics perspective

Though acceleration is more directly informative about physical impulses onto the body and was the most reliable kinematic parameter associated with acoustics in the current study, we are cautious in interpreting this as direct evidence for the gesture–speech physics hypothesis as it has been originally proposed.[22–24,68] In this study, we only observe associations and we have not systematically varied kinematic and acoustic conditions so as to probe putative causal relations between forces and acoustics. Additionally, we obtain that performers' vocal peaks *followed* manual acceleration peaks but *preceded* manual

deceleration peaks. This suggests that the connection between force transfer and vocal movements in this context is not straightforward (as would be the case if all moments of force transfer connected with acoustic peaks), but rather that gestural movements are managed by performers to take into account both aesthetic and biophysical constraints. Further, it should be emphasized that acceleration is a kinematic variable and it might be that the *performance variable* for the performer is also kinematic in nature; for example, the vocalizer might aim to change the direction of the movement during a vocal inflection to visually signal to the audience. Sudden changes in direction will also be preceded and followed by negative and positive peaks in the acceleration profile of the gesture. Further, head gesture accelerations have been found to align with external beats,[69] and these head gestures seem to function to *visually* signal piece onsets and tempo.[70] Thus, acceleration as a kinematic variable that couples most reliably with some kind of structure in sound is a finding open to multiple interpretations regarding the underlying mechanisms.[71,72]

At the same time, manual accelerations will by necessity involve force transfers in the musculoskeletal system. Given that there is accumulating evidence that gestures can affect the voice directly through force transfer via the respiratory system (as evidenced by studies relating mass of the movement articulators and acceleration, to chest kinematics and vocalization),[73] and given that the current study is also about voicing and gesturing and shows highly comparable relationships in a music context, we conclude and emphasize that the current findings are in line with the general gesture–speech physics thesis.

Gesture–vocal coupling did diverge in some subtle ways from what has been observed previously in steady phonation and fluent speech. For example, in phonation and speech, the perturbing effects of gestures seem to have more extreme effects on the amplitude envelope than F0,[22,23] and here, we obtain an opposite pattern, where F0 is most related to gesture. However, there are important differences between this study and previous research on gesture–speech physics in which research subjects were instructed to *inhibit* effects on speech-vocalization due to upper limb movement. Given that laryngeal counter adjustments can be flexibly employed in rapid response to sudden increases in subglottal pressure (which might be more difficult for respiratory muscle adjustments controlling amplitude), the F0 effects due to upper limb movement diminish relative to effects on the amplitude of the vocalization. Interestingly, if we apply this reasoning to the Karnatak music context, it is quite possible that the physical effects of upper limb movements on vocal acoustics have become ritualized to some extent as they have, over time, been found to affect performance in ways that can strategically be brought into alignment with the aesthetic target. This would explain why we find that F0 is more coupled to kinematics, as opposed to vocal amplitude, which is more directly related to respiratory changes. Thus, F0 and gesture coupling might reflect a cooperation of vocal cords tensioning with the effects on the respiratory system by manual gesture–speech physics. Note that this moves us toward an understanding that gesture–vocal coupling does not arise *purely* out of biomechanics, as if there is no brain or other situatedness required. Not bodies, not brains, not the environment, *but persons* perform actions, and these actions are under a multitude of biomechanical, cultural, and neural constraints,[74,75] which can be brought into productive harmony or resonance.[76] Concretely, this means that culturally typical gestures occurring within this musical context will have evolved in a way that does not oppose the performers' respiratory-vocal actions. Rather, we suggest that the gestures develop, both in the long and short term (across lineages and within the bodily experience of each individual vocalist) to be in physical harmony or resonance with the performers' respiratory-vocal actions. The gestures are accordingly constrained by biophysics *and* the cultural-aesthetic goals of the performer—hence the entanglement.

Gesture–vocal coupling and bodily tensegrity: an aesthetic entanglement

In our theoretical contribution to this article, we propose that gesture–vocal coupling should be considered in relation to the tensegrity structure of the body, as well as within the wider aesthetic and performance context. In this sense, it can be viewed as a neural–bodily distributed aesthetic entanglement. This entanglement occurs within the context of kinetic cross-modal mappings that are best understood as an active sensing and perturbing of the deformations of a prestressed tensegrity-structured body due to gesture-induced physical impulses. Tensegrity structures involve tension (connective tissues and muscles) and compressive elements (bones) that form a networked architecture. This imbues such systems with particular dispositions,[77] which are characteristics of many living systems, such as (human-) animal bodies, as well as cells. One of these dispositions is that there is always some level of tension, which naturally distributes locally induced forces over more peripherally connected sets of muscoskeletal elements.[78] This pre-stress entails that tensioning one element (left hand fist clenching) can affect movement parameters (e.g., stiffness and amplitude) of more peripheral elements (right arm movement).[79] In the case of vocalizing and moving the upper limbs, we argue that the tensegrity structure of the body creates the conditions wherein a forceful gesture can affect respiratory-vocal processes as the cascading effects of moving one body part can affect respiration and, therefore, vocalization.[22] It is not just that a bodily action can impact the wider muscoskeletal system "downstream" in this way, but also that the perception of those effects[80] becomes part of the aesthetic performance itself. That is, the "dynamical causal loop"[81] between gestural action and the sensed constraints on the respiratory-vocal system is regulated in relation to the musical context in each vocal performance.

We describe gesture–vocal coupling as an aesthetic entanglement in the sense that gesture-induced physics are brought in harmony with the performer' goals, linked to aesthetic norms that are part of the musical practice. The gesture–vocal complex is thus distributed across the neural-bodily system as influenced by factors, such as the particular body and tensegrity structure of the performer,[82] cross-modal perception, the performer's personal history (e.g., the influence of their teacher and learning process), and the structure and character of the performance. All of this is brought into active neural–bodily harmony through gesturing and vocalizing in biomechanically stable ways.

Future directions

This study's findings regarding acceleration being the best statistical predictor for gestural–vocal coupling are consistent with the interpretation that mappings between force-producing movements and acoustic change are salient for gesture–vocal coupling in this context. This is consistent with observations by Paschalidou regarding the significance of gestural enactments of effort in North Indian vocal performance.[5] However, the findings of the current study do not prove the interpretation of force transfer being salient. We suggest that future research should, therefore, target this specific mapping in co-singing and co-musicking gesturing. An important future research endeavor is a more fine-grained study of which gesture-related muscle groups, implicated in respiratory control, are best functionally recruited during which vocal targets. Further, bodily gestures that can interact with the vocal system need not be limited to hand gestures,[83–85] and indeed, it is apparent in the current performances that movements were performed with the whole upper body. Given that, for example, postural changes associated with piano-playing seem to affect superior airflow, affecting the harmonic formant of vocal acoustics when simultaneously singing and playing the piano,[86] it is possible that other types of bodily postures and movements are entangled with the voice. This more fine-grained whole-body research would employ respiratory-related measurements, as well as muscle activation tracking (EMG) in experimental contexts, where different vocal targets need to be reached, with and without specific gestures designed to recruit different muscle units. It is quite possible that vocalists can reach vocal targets with and without gesture equally well, but that they use different coordinative muscle units in each case to reach such vocal targets.[20,21] It is further possible that specific muscles related to inspiratory versus expiratory modulations are more likely to be recruited in gesture for certain vocal inflections (F0 descent) rather than others (F0 rise). Thus, we think the current research reveals just the tip of the iceberg in terms of the actual sensorimotor solutions that are reached in these and many other gesture–vocal coupling practices.

In addition, more research is needed that addresses the generalizability of particular gesture–vocal coupling across and within performance styles. Within styles, it is important to note that the number of professional performers who participated in this study constitutes a relatively low "sample size" if weighted against conventional standards in psychology, and thus, we should caution to make any sweeping generalizations from the current data alone. Note that the number of samples taken for this study is high and contributes to the reliability of the current results—there were many gesture events (n = 1630) collected, which were recorded during multiple performances per performer (total performances, n = 35). Across styles, we would hypothesize that biomechanical stabilities between vocalization and gesture would apply to all gesture–vocal performances. However, the ways in which this is manifest are likely to vary depending on the aesthetic qualities admired in the style. For example, much contemporary opera performance tends to favor relatively naturalistic acting, without additional gesturing. Therefore, it would be important in each case to examine how the biomechanical constraints discussed here are brought in harmony with the aesthetics and sociocultural context of the particular style.

Finally, viewing gesture–vocal coupling as a neural–bodily distributed aesthetic entanglement invites systematic research into connections between musical syntax (ragas, phrases, and motifs) and gestural physics, both within and across performers. This aesthetic entanglement perspective considers the harmony between the physical tensegrity structure of human bodies and the aesthetic goals of the specific music performance. We propose that through such an approach, progress can be made in understanding why gesture manifests as it does in musical contexts.

ACKNOWLEDGMENTS

We would like to thank the Karnatak vocalists, Akkarai Subhalakshmi, Pattabhirama Pandit, Hemmige Prashanth, and Brindha Manickavasakan, who performed for this study. Many thanks also to Rainer Polak for his significant assistance in recording the audiovisual and motion capture data on which this study is based, and to Nikita Kudakov for his work in logging the data. Finally, we would like to thank two anonymous reviewers and the nonanonymous reviewer (Ramesh Balasubramaniam) for their helpful comments for improving this paper. This research has been cofunded by a VENI grant (VI.Veni.201G.047) awarded by the Dutch Research Council (NWO) to Wim Pouw (PI).

Open access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Each author contributed equally to the experimental and writing requirements of the manuscript.

OPEN DATA AND ANALYSIS SCRIPT

All deidentified data are available at the Open Science Framework (https://osf.io/ux48y/).

PEER REVIEW

The peer review history for this article is available at: https://publons.com/publon/10.1111/nyas.14806.

ORCID

Lara Pearson https://orcid.org/0000-0002-5073-8738
Wim Pouw https://orcid.org/0000-0003-2729-6502

REFERENCES

1. Davidson, J. W. (2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae*, *5*(2), 235–256.
2. Clayton, M. (2007). Time, gesture and attention in a Khyāl performance. *Asian Music*, *38*(2), 71–96.
3. Rahaim, M. (2012). *Musicking bodies: Gesture and voice in Hindustani music*. Wesleyan University Press.
4. Pearson, L. (2016). *Gesture in Karnatak music: Pedagogy and musical structure in South India*. Durham University.

5. Paschalidou, P.-S. (2017). *Effort in gestural interactions with imaginary objects in Hindustani Dhrupad vocal music*. Durham University.
6. Leante, L. (2009). The lotus and the king: Imagery, gesture and meaning in a Hindustani Rāg. *Ethnomusicology Forum*, *18*(2), 185–206.
7. Mani, C. (2017). Gesture in musical declamation: An intercultural approach. *Musicologist*, *1*(1), 6–31.
8. Clarke, E. (2001). Meaning and the specification of motion in music. *Musicae Scientiae*, *5*(2), 213–234.
9. Godøy, R. I. (2010). Gestural affordances of musical sound. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 103–125). Routledge.
10. Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance* (pp. 55–83). Cambridge University Press.
11. Windsor, W. L. (2011). Gestures in music-making: Action, information and perception. In A. Gritten & E. King (Eds.), *New perspectives on music and gesture* (pp. 45–66). Ashgate Publishing.
12. Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, *5*(1), 685–697.
13. Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89.
14. Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. *Proceedings of the International Congress of Phonetic Sciences*, *2019*, 5.
15. Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232.
16. McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research*, *27*(1), 69–89.
17. Ebert, D., Raßler, B., & Hefter, H. (2000). Coordination between breathing and forearm movements during sinusoidal tracking. *European Journal of Applied Physiology*, *81*(4), 288–296.
18. Hodges, P. W., & Gandevia, S. C. (2000). Activation of the human diaphragm during a repetitive postural task. *Journal of Physiology*, *522*(Pt 1), 165–175.
19. Hodges, P. W., & Gandevia, S. C. (2000). Changes in intra-abdominal pressure during postural and respiratory activation of the human diaphragm. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *89*(3), 967–976.
20. Cooper, B. G., & Goller, F. (2004). Multimodal signals: Enhancement and constraint of song motor patterns by visual display. *Science*, *303*(5657), 544–546.
21. Lancaster, W. C., Henson, O. W., & Keating, A. W. (1995). Respiratory muscle activity in relation to vocalization in flying bats. *Journal of Experimental Biology*, *198*(Pt 1), 175–191.
22. Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2020). Energy flows in gesture–speech physics: The respiratory-vocal system and its coupling with hand gestures. *Journal of the Acoustical Society of America*, *148*(3), 1231–1247.
23. Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2020). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, *1491*(1), 89–105.
24. Pouw, W., Harrison, S. J., & Dixon, J. A. (2019). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391–404.
25. Cavallari, P., Bolzoni, F., Bruttini, C., & Esposti, R. (2016). The organization and control of intra-limb anticipatory postural adjustments and their role in movement performance. *Frontiers in Human Neuroscience*, *10*, 525.
26. Levin, S. M. (1997). Putting the shoulder to the wheel: A new biomechanical model for the shoulder girdle. *Biomedical Sciences Instrumentation*, *33*, 412–417.
27. Cordo, P. J., & Nashner, L. M. (1982). Properties of postural adjustments associated with rapid arm movements. *Journal of Neurophysiology*, *47*(2), 287–302.
28. Hodges, P. W., & Richardson, C. A. (1997). Feedforward contraction of transversus abdominis is not influenced by the direction of arm movement. *Experimental Brain Research*, *114*(2), 362–370.
29. Ferstl, Y., Neff, M., & McDonnell, R. (2020). Understanding the predictability of gesture parameters from speech and their perceptual importance. Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, 1–8.
30. Fatone, G. (2010). "You'll break your heart trying to play it like you sing it": Intermodal imagery and the transmission of Scottish classical bagpiping. *Ethnomusicology*, *54*(3), 395–424.
31. Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception*, *23*(3), 221–247.
32. Eitan, Z., & Tubul, N. (2010). Musical parameters and children's images of motion. *Musicae Scientiae*, *Special Issue*, *14*(Issue 2_suppl) 89–111.
33. Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, *5*, 789.
34. Nymoen, K., Godøy, R. I., Jensenius, A. R., & Torresen, J. (2013). Analyzing correspondence between sound objects and body motion. *ACM Transactions on Applied Perception*, *10*(2), 1–22.
35. Nymoen, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analyzing sound tracings: A multimodal approach to music information retrieval. Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies, 39–44.
36. 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.
37. Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology-Human Perception and Performance*, *35*(6), 1791–1810.
38. Bosker, H. R., & Peeters, D. (2021). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B*, *288*, 20202419.
39. Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995.
40. 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.
41. Gaver, W. W. (1993). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, *5*(4), 285–313.
42. Vanderveer, N. J. (1980). Ecological acoustics: Human perception of environmental sounds. *Dissertation Abstracts International*, *40*(9-B), 4543.
43. Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704.
44. Fatone, G., Clayton, M., Leante, L., & Rahaim, M. (2011). Imagery, melody and gesture in cross-cultural perspective. In A. Gritten & E. King (Eds.), *New perspectives on music and gesture* (pp. 203–220). Ashgate Publishing.
45. Pearson, L. (2022). Inscription, gesture and social relations: Notation in Karnatak music. In E. Payne & F. Schuiling (Eds.), *Material cultures of music notation: New perspectives on musical inscription* (pp. 139–151). Routledge.
46. Krishna, T. M., & Ishwar, V. (2012). *Carnatic music: Svara, Gamaka, Motif and Raga identity*. http://repositori.upf.edu/handle/10230/20494
47. Pearson, L. (2016). Coarticulation and gesture: An analysis of melodic movement in South Indian raga performance. *Music Analysis*, *35*(3), 280–313.
48. Viswanathan, T. (1977). The analysis of rāga ālāpana in South Indian music. *Asian Music*, *9*(1), 13–71.
49. Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, *41*(3), 841–849.

50. Kita, S., Van Gijn, I., & Van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and sign language in human–computer interaction, International Gesture Workshop Bielefeld, Germany, September 17–19, 1997, Proceedings. Lecture Notes in Artificial Intelligence* (pp. 23–35). Springer-Verlag.
51. Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer (Version 6.1. 05)[Computer application]*. https://www.praat.org.
52. He, L., & Dellwo, V. (2017). Amplitude envelope kinematics of speech signal: Parameter extraction and applications. *Journal of the Acoustical Society of America*, *141*(5), 3582–3582.
53. Hastie, T. (2022). *gam: Generalized additive models* (1.20.1) [Computer software]. https://CRAN.R-project.org/package=gam
54. Aruin, A. S., & Latash, M. L. (1995). Directional specificity of postural muscles in feed-forward postural reactions during fast voluntary arm movements. *Experimental Brain Research*, *103*(2), 323–332.
55. Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *Journal of the Acoustical Society of America*, *65*(5), 1271–1275.
56. Borchers, H. W. (2022). *pracma: Practical numerical math functions* (2.3.8) [Computer software]. https://CRAN.R-project.org/package=pracma
57. Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling. *Journal of Phonetics*, *70*, 86–116.
58. Clayton, M., Jakubowski, K., & Eerola, T. (2019). Interpersonal entrainment in Indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure. *Musicae Scientiae*, *23*(3), 304–331.
59. Jakubowski, K., Eerola, T., Alborno, P., Volpe, G., Camurri, A., & Clayton, M. (2017). Extracting coarse body movements from video in music performance: A comparison of automated computer vision techniques with motion capture data. *Frontiers in Digital Humanities*, *4*, *9*.
60. Eerola, T., Jakubowski, K., Moran, N., Keller, P. E., & Clayton, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science*, *5*, 171520.
61. Zelic, G., Kim, J., & Davis, C. (2015). Articulatory constraints on spontaneous entrainment between speech and manual gesture. *Human Movement Science*, *42*, 232–245.
62. Cliff, O. M., Lizier, J. T., Tsuchiya, N., & Fulcher, B. D. (2022). Unifying pairwise interactions in complex dynamics. ArXiv:2201.11941 [Physics]. http://arxiv.org/abs/2201.11941
63. Kuhn, M. (2022). *caret: Classification and regression training* (6.0-92) [Computer software]. https://CRAN.R-project.org/package=caret
64. Ćwiek, A., & Fuchs, S. (2021). *Hand–mouth coordination in a pointing task requiring manual precision*. https://issp2020.yale.edu/ProcISSP2020.pdf
65. Graham, K. M., Moore, K. D., Cabel, D. W., Gribble, P. L., Cisek, P., & Scott, S. H. (2003). Kinematics and kinetics of multijoint reaching in nonhuman primates. *Journal of Neurophysiology*, *89*(5), 2667–2677.
66. Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, *7*(2), 431–437.
67. Perrier, P., & Fuchs, S. (2008). Speed–curvature relations in speech production challenge the 1/3 power law. *Journal of Neurophysiology*, *100*(3), 1171–1183.
68. Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, *117*(21), 11364–11367.
69. Bishop, L., & Goebl, W. (2018). Communication for coordination: Gesture kinematics and conventionality affect synchronization success in piano duos. *Psychological Research*, *82*(6), 1177–1194.
70. Bishop, L., & Goebl, W. (2018). Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo. *Psychology of Music*, *46*(1), 84–106.
71. Luck, G., & Sloboda, J. (2008). Exploring the spatio-temporal properties of simple conducting gestures using a synchronization task. *Music Perception*, *25*(3), 225–239.
72. Luck, G., & Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception*, *26*(5), 465–473.
73. Pouw, W., & Fuchs, S. (2021). *Origins of vocal-entangled gesture*. https://doi.org/10.31234/osf.io/egnar.
74. Damm, L., Varoqui, D., De Cock, V. C., Bella, S. D., & Bardy, B. (2020). Why do we move to the beat? A multi-scale approach, from physical principles to brain dynamics. *Neuroscience & Biobehavioral Reviews*, *112*, 553–584.
75. Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B*, *376*(1835), 20200334.
76. Raja, V. (2020). Resonance and radical embodiment. *Synthese*, *199*(Suppl 1), S113–S141.
77. Turvey, M. T., & Fonseca, S. T. (2014). The medium of haptic perception: A tensegrity hypothesis. *Journal of Motor Behavior*, *46*(3), 143–187.
78. Carello, C., Silva, P. L., Kinsella-Shaw, J. M., & Turvey, M. T. (2008). Muscle-based perception: Theory, research and implications for rehabilitation. *Brazilian Journal of Physical Therapy*, *12*(5), 339–350.
79. Silva, P., Moreno, M., Mancini, M., Fonseca, S., & Turvey, M. T. (2007). Steady-state stress at one hand magnifies the amplitude, stiffness, and non-linearity of oscillatory behavior at the other hand. *Neuroscience Letters*, *429*(1), 64–68.
80. Jékely, G., Godfrey-Smith, P., & Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1821), 20190764.
81. Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.
82. Caldeira, P., Davids, K., & Araújo, D. (2021). Neurobiological tensegrity: The basis for understanding inter-individual variations in task performance. *Human Movement Science*, *79*, 102862.
83. Miller, N. A., Gregory, J. S., Aspden, R. M., Stollery, P. J., & Gilbert, F. J. (2014). Using active shape modeling based on MRI to study morphologic and pitch-related functional changes affecting vocal structures and the airway. *Journal of Voice*, *28*(5), 554–564.
84. Pettersen, V. (2006). Preliminary findings on the classical singer's use of the pectoralis major muscle. *Folia Phoniatrica et Logopaedica*, *58*(6), 427–439.
85. Pettersen, V., & Westgaard, R. H. (2005). The activity patterns of neck muscles in professional classical singing. *Journal of Voice*, *19*(2), 238–251.
86. Longo, L., Di Stadio, A., Ralli, M., Marinucci, I., Ruoppolo, G., Dipietro, L., de Vincentiis, M., & Greco, A. (2020). Voice parameter changes in professional musician-singers singing with and without an instrument: The effect of body posture. *Folia Phoniatrica et Logopaedica*, *72*(4), 309–315.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pearson, L., & Pouw, W. (2022). Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement. *Ann NY Acad Sci.*, *1515*, 219–236. https://doi.org/10.1111/nyas.14806