ARTICLE

Check for updates

Hidden musicality in Chinese *Xiangsheng*: a response to the call for interdisciplinary research in studying speech and song

Francesca R. Sborgi Lawson [1✉]

Recent scholarship in the field of music cognition suggests the need for increased inter-disciplinarity in moving beyond the boundaries of Western European music, and the study of the relationship between language and music is an especially fruitful area that can benefit from interdisciplinary collaboration. This paper heeds the call for collaborative research in cultures outside of Western Europe by focusing on *Xiangsheng*, a form of Chinese musical comedy that features an intriguing relationship between speech and song in performance. The present paper argues that analytical tools and perspectives from conversational analysis, communicative musicality, empirical research on music-language relationships, and performative mutuality in ethnomusicology all speak to the idea of *musicality*—the underlying capacity that undergirds our ability to communicate both verbally and musically—as a common foundational behavior in both speech and song. Musicality is particularly apparent in the way Chinese *Xiangcheng* actors relate to each other and their audiences in both spoken and musical modalities, and this paper suggests how judiciously employing a variety of methodological approaches in the study of musicality can yield important insights for researchers from all disciplines.

[1] Brigham Young University, Provo, UT, USA. ✉email: Francesca_Lawson@byu.edu

Introduction

In October 2018, a number of scholars gathered at the Max Planck Institute for Empirical Esthetics in Frankfurt, Germany to discuss the future of cross-cultural work in the field of music cognition (Jacoby et al., 2020). The scholars agreed that the vast majority of the research conducted in music psychology has heretofore involved Western participants and focused primarily on Western music, so they recommended that future research in music psychology should consider (1) increased interdisciplinary collaboration to promote cross-cultural work and (2) a new emphasis on ways to overcome disciplinary differences in assumptions, methods, and terminology, particularly with regard to the distinctions between empirical and musicological approaches (p. 185).

Certainly, examples of interdisciplinary research projects that focus on music already abound. For example, Savage (2019) and Cross et al. (2013) consider cultural evolutionary perspectives in music research, and there are several excellent studies initiated by music psychologists and neuroscientists that speak to issues that involve cross-cultural questions, such as Margulis et al. (2019) and Mehr et al. (2019). Although somewhat less prevalent, there are also studies initiated by ethnomusicologists, many of whom focus on music outside of the Western European tradition, while collaborating with or employing methods from scientific disciplines, including Becker (2009), Tolbert (2012), Widdess et al. (2019), and Fatone et al. (2011). Despite these important interdisciplinary studies, however, the participants at the Frankfurt meeting remind us that continued scholarly collaboration is necessary for cross-cultural research to survive (Jacoby et al., 2020, pp. 185–186).

As a response to this call, this paper presents a case study of Chinese musical comedy that is rooted in ethnomusicological fieldwork—an example from outside the Western European musical tradition that encourages a broadening of intercultural perspectives. Further, the current study focuses on *musicality*, which is the underlying foundational capacity that undergirds our ability to communicate both verbally and musically. Since the relationship between music and language is an area of research that Aniruddh Patel claims is of particular interest for both scientists and ethnomusicologists in the study of music cognition (2008, p. 417), a focus on musicality is potentially significant to scholars from a variety of different fields. Relying on statistical, linguistic, and ethnomusicological methods, the research described in this paper also represents a collaboration that bases many of its premises on concepts and methodologies from several unexpected and unrelated disciplines, including sociology, developmental psychology, music psychology and neuroscience. The following section discusses some of the different fields that have contributed to the study of the underlying musical elements of language—elements that are critical to understanding how speech and song are related in Chinese musical comedy.

Musical aspects of speech in the study of intersubjective communication

The subtle ways in which people move, talk, and gesture have captivated researchers from a variety of disciplines. Based on the work of scholars, such as Gordon Allport and Vernon (1933), Erving Goffman (1979), Nalini Ambady (1992), and Ann Wennerstrom (2001), we have learned how expressive verbal and non-verbal behaviors are especially rich in social information (Ambady, p. 256) and demonstrate nuanced, highly synchronized, and often unconscious rhythms of conversation and bodily movements—aspects of speech that appear to be musical (Schegloff, 2007; Sidnell, 2010; Sidnell and Stivers, 2012). Until recently, however, music scholars have generally not contributed substantively to the mainstream discussion about these expressive forms of communication, and the word "music" has rarely, if ever, been used by scholars who study expressive behavior in adults[1].

In addition to the research on verbal and non-verbal communication in adult interactions, child psychologists, sociolinguists, and pediatricians have been studying the highly sophisticated exchanges between mothers and their pre-linguistic infants for the past six or more decades. Using similar methods of videotaping interactions as their counterparts who focus on the expressive behavior of adults, scholars studying mother-infant interactions have made remarkable discoveries about the delicate ways infants and their mothers relate to one another. Detailed microanalyses of videotaped interactions between mother and infant demonstrate how the mother's exaggerated, rhythmic, singsong vocalizations are directed toward and, even more importantly, solicited by the infants. Stephen Malloch and Colwyn Trevarthen (2010) eventually coined the term "communicative musicality" (CM), which has heretofore been called "motherese," to express the musical and dance-like nature of the proto-conversations between infants and their mothers. They believe that CM reflects a human ability to share a sense of time, shape jointly created pitch contours, and move with anticipated rhythms and emotions (Trevarthen and Malloch, 2000, p. 3). Believing that it is essential to acknowledge the *musicality* inherent in the bodily and vocal expression used in managing human relationships, they argue that communicative musicality is an appropriate and descriptive term to depict these rhythmic, melodic, and kinetic gestures (3). Their use of the word musicality points to the human abilities that are not music, per se, but instead represent the *capacity* (1) to speak, make music, dance, and engage in all the other temporal arts and (2) to interact with others in a communicatively intimate way. They explain:

> We define musicality as expression of our human desire for cultural learning, our innate skill for moving, remembering and planning in sympathy with others that makes our appreciation and production of an endless variety of dramatic temporal narratives possible—whether those narratives consist of specific cultural forms of music, dance, poetry or ceremony; whether they are the universal narratives of a mother and her baby quietly conversing with one another; whether is it the wordless emotional and motivational narrative that sits beneath a conversation between two or more adults or between a teacher and a class… It is our common musicality that makes it possible for us to share time meaningfully together… (4–5).

Ellen Dissanayake contributes to the argument about CM by providing an ethological explanation for its centrality in human evolution. She argues that the emergence of bipedality necessitated a narrowing of the hips and a reshaping of the mother's pelvis, which reduced fatigue during upright locomotion but resulted in giving birth to progressively large-brained babies through an increasingly narrowed birth canal (Dissanayake, 2010, p. 22; Malotki and Dissanayake, 2018, p. 202). As a consequence of these anatomical changes, the gestation period was reduced, requiring constant, attentive care from adults for much longer than any other primate (Malotki and Dissanyake, 2018, p. 202). The behavioral adaptation that arose from the anatomical changes was the communicative behavior that would assure intense maternal care for the helpless infant—what Malloch and Trevarthen refer to as communicative musicality (Dissanayake, 2010, p. 22). Differing from the mode of conversation used with other adults and older children, communicative musicality is characterized by its "higher overall tone, wider tone range, slower

tempo, exaggerated vowels, repetitions, and a simplified, specialized vocabulary" (Malotki and Dissanayake, 2018, p. 203). Dissanayake also agrees that musicality is an appropriate label since the interactions employ melodic vocal contours, are organized in rhythmic bouts over time, and utilize expressive dynamic contrasts and variations, all the while interspersed with periods of silence between bouts. Moreover, the interactions are multimodal for both mother and infant since vocal, facial, and bodily movements occur together according to regular rhythmic pulses (Malotki and Dissanayake, 2018, p. 203).

Musicality and music

Scholars who study mother-infant communication are not the only ones to recognize the significance of musicality in human communication. Henkjan Honing (2018, p. 3) explains that musicality is a predisposition for music (as well as language)—a capacity that becomes clear when the focus is on perception rather than production. Significantly, Honing's recently edited volume is entitled *The Origins of Musicality*, with the contributors focusing on the biological capability to be musical rather than on music per se. Distinguishing between musicality and music is central in understanding the theoretical mindset of scholars in the sciences, many of whom focus on the biological underpinnings of musicality, compared to scholars in the humanities, most of whom focus on cultural differences in creating, performing, and consuming music.

Despite some of the obvious differences between the musicality that undergirds music and the musicality that is inherent in the expressive behavior studied by linguists, sociologists, and researchers who study mother-infant communication and linguists, might there also be commonalities? I argue that finding areas of congruence between the musicality of speech and song may be a fruitful endeavor that could shed light on both music and language, allowing scholars from different disciplinary backgrounds to provoke each other into insights that neither would have found independently. An example of how one might pursue such an interdisciplinary question comes from the following invitation by Dissanayake.

Three points that beg further investigation. Dissanayake suggests three intertwined points that invite further investigation in CM: (1) the noteworthy nature of the signals presented by the mother; (2) the infant's strong and untaught receptivity to the signals; and (3) the infant's active contribution to the communication (2010, p. 23). Beyond the field of mother-infant research, music scholars who study the interaction between performers and audience members—an interaction that embodies many of the aspects of CM—might also accept Dissanayake's invitation to look more carefully at the relationship between a presented signal, its reception, and the receiver's contribution back to the signaler in musical performance. Since Thomas Turino's concepts of "presentation" and "participation" (2008) have been more commonly used in the ethnomusicological literature I will use those two terms to refer to what Dissanayake calls the "signal" and the "contribution," respectively. Moreover, because presentation and participation are more easily observed and measured than the more elusive aspect of reception, I will focus primarily on presentation and participation and only secondarily on how the notion of reception, as Dissanayake's second point, may present an opportunity for future research in ethnomusicology.

Communicative musicality in ethnomusicology. One could certainly argue that music scholars have already addressed similar questions about communicative interaction among performers. For example, studies of jazz performance have contributed

substantially to an understanding of the intimate relationship between performers who present or create the initial signal and other performers who respond and contribute back to the signaler. In *Thinking in Jazz* (1994), Paul Berliner details the many ways jazz performers respond musically to each other by matching timbres, adjusting beat placement, mirroring rhythmic ideas, encouraging soloists, and engaging in imitative interplay (pp. 346–371). He explains that this kind of nuanced interaction requires what one musician called "dividing your senses" while listening to multiple band members at once (p. 362).

Additionally, jazz musicians converse musically with members of the audience. As early as 1951, sociologist Howard Becker noted that audiences at jazz concerts were involved with performances in ways that were "much more than casual" (1951, p. 136). Audience responses, verbal or otherwise, cause the musicians to change the way they play, creating what Berliner calls a "communication loop" (1994, p. 459). Musicians listen carefully to how the audience responds, and if the response is positive, the performer develops the musical idea further, thereby continuing to excite the crowd. As Charles Hersch explains, "This mutability of the musical self leads musicians to play things in the context of a group that they have never played before and are as surprised as anyone to hear them emanating from their instrument" (2019, p. 371). Of course, jazz performances are not always conversational; however, when spontaneous banter does occur between musicians and between musicians and their audiences, that performance becomes a highly charged event (Hamilton, 2007, p. 114, 199).

In addition to jazz scholarship, evidence of reciprocity between performers and audience members abounds in ethnomusicological research (Middleton, 1990; Small, 1998; Frith, 1998; Hesmondalgh and Negus, 2002; Nettl, 2005; Turino, 2008; Tsioulakis and Hytonen-Ng, 2017). For example, the various case studies contained in the edited volume *Musicians and their Audiences* feature an impressive array of scholarship on performative mutuality. From Bruce Johnson's application of cognitive theory to observations of every-day music (2017) to Elina Hytonen-Ng's description of the way jazz musicians articulate their expectations of audience members (2017), this volume provides a treasure trove of materials and methodologies to study the fascinating complexities of performer-audience interaction. Editors Ioannis Tsioulakis and Elina Hytonen-Ng underscore the notion that audiences are as intrinsic to music making as performers—an idea that has "achieved consensus in culturally/socially based musicological writings" (2017, p. 1). Their central thesis is that the musician-audience relationship should be seen on a continuum of interactive possibilities rather than as discrete, dichotomous modes of interaction, and researchers should focus on the dynamic nature of the relationship between audiences and performers through an ongoing practice of renegotiation. In the end, Tsioulakis and Hytonen-Ng argue that the division between performers and spectators is not as stable and self-explanatory as musicologists might have once thought (2017, p. 12), and they refer to the dynamic reciprocity inherent in the performer-audience relationship as "performative mutuality" (p. 6).

Guiding many of the discussions in the volume is Thomas Turino's distinction between participatory performances, in which there is no distinction between artist and audience, and presentational performances, in which a group of artists provides music for an audience that does not participate in making the music or in dancing (2008, pp. 26–65). However, many of the contributors also challenge and problematize Turino's dichotomy. For example, Laura Leante acknowledges that Hindustani classical musicians engage in a presentational performance (2017, p. 34), yet she proposes an interpretation of the

3

performance event that recognizes the importance of audience participation, albeit not quite on the level described by Turino. By using one camera that provided a frontal shot of the stage and another one focused on the audience (2017, pp. 39–40), Leante was able to see an increase in the clapping of the *tal* during the drummer's solo, giving the audience a tool to keep their bearings. In doing so, Leante argues that people with different abilities can join and contribute to the intensity of the performance—a situation actively encouraged by the vocal soloist, who considers this interaction as part of a strategy in creating a positive reception of the music by the audience (p. 46). While still presentational, Leante notes that there are instances in which audiences become an active part of music production through a sense of *communitas* (p. 44), a term originally recast by Edith Turner that refers to "the relation quality of full unmediated communication, even communion, between definite and determinate identities, which arises spontaneously" (Turner, 1977, p. 46).

Even in presentational settings where the audience is sitting quietly throughout a performance, audience members can still be deeply involved in a more muted—though no less intense—kind of connection. For example, in the film, *Singing the Dark Away* (Davitt, 1996), a documentary about Joe Heaney, the late *sean-nos* master, a short snippet of a video recording of a *sean-nos* singer performing for a small group of listeners provides a vivid example of a highly engaged audience in a presentational performance. From minute 9:40–10:52 we see the *sean-nos* singer holding the hand of one of the male listeners directly in front of him, moving his arm clockwise in a small circle, as though the singer is transmitting the song directly to the listener. Although the performance is presentational, this male listener, in addition to the others who are listening but not moving their hands, epitomizes what Ruth Herbert refers to as an example of highly engaged absorption (2011, p. 50)—an intense listening response that does not display the vigorous bodily movement and vocal responses characteristic of Turino's definition of a participatory performance, but is nevertheless participatory in a different way.

As absorption is difficult to observe and even harder to quantify, ethnomusicologists have struggled in discussing this aspect of audience engagement, often failing to explain the nuances of absorptive listening and sometimes even misreading and misinterpreting responses that do not fit neatly into the presentational-participatory paradigm. Consequently, I suggest that we take a cue from researchers in expressive behavior and CM to see how we might use technology to expand our understanding of the complex reciprocal relationships between audience members and performers. The goal of expanding our methodological approach would be to gain a more detailed understanding of the symbiotic nature of performative mutuality, thereby allowing us to enter into a conversation with other scholars who engage in similar kinds of intersubjective research in other fields.

The Cambridge Study
Research published in 2016 from the University of Cambridge provides an example of a study that can further such interdisciplinarity on the topic of intersubjective communication. The Cambridge researchers used ELAN, a video analysis software, to gather and analyze data regarding pulse and intonation in the spontaneous spoken interactions between eight pairs of same-sex friends in video-recorded performances (Hawkins et al., 2013; Robledo et al., 2016; Cross et al., 2016). What makes this study unusual is that the researchers discovered that the spoken interactions between the friends became both rhythmic and pitched when the friends were highly engaged with each other; the

researchers referred to these emotionally engaged interactions as "successful" when the two friends were "attitudinally aligned." In other words, the researchers discovered regular rhythmic cycles when the first syllable of one speaker arrived on beat with the pulse established by the previous speaker in the spoken interactions between the friends. Additionally, their results also suggested that these regular rhythmic cycles were sometimes accompanied by the systematic use of pitch intervals between the final accent of the first speaker's utterance and the initial pitch of the second speaker's response (Robledo et al., 2016). The authors concluded that speech and song may both be underpinned by common neurological processes in certain contexts when the speakers are emotionally engaged (or attitudinally aligned, as they say in the study)—a finding that supports Cross's previous claim that speech and music may be seen as two halves of the human communicative toolkit (2009, 2014).

Speech, music, and performative mutuality in *Xiangsheng*. After learning about the Cambridge study, I immediately thought about the conversational dialog and musical exchanges between the two actors featured in the Chinese narrative genre known as *xiangsheng* (XS) or crosstalk, a genre I researched during the course of my fieldwork (Lawson, 2011, pp. 113–124; Lawson, 2017, pp. 88–106). XS is an interesting example in light of the Cambridge study because it is a performance tradition that displays the porous border between speech and song during the course of performance. I wondered if using ELAN to analyze a XS performance might demonstrate how findings from an empirical study might ultimately drive a research project on intersubjectivity—specifically looking at the primary actor's presentational signal and the second actor's contribution in a real-world performance.

Perry Link (1980, p. 84) translates XS literally as "face and voice routines"—an interesting moniker when considering the multimodal nature of dialog, whether it be between Chinese actors, same-sex friends as in the Cambridge study, or a mother and her infant. The two Chinese actors, known respectively as the "joke cracker" (*dougende*) and the "joke setter" or straight man (*penggende*) (Moser, 1990, p. 46)), participate in a comedic dialog that appears to be improvised but is actually written by an author who specializes in the genre. Although the ability to internalize a script is a requirement for an actor in any (literary-centric) culture, the slapstick nature of XS makes a spontaneous performance especially difficult for all but the best actors (Lawson, 2011, pp. 115–116).

This form of scripted comedy is conceived in four sections, with the first three building up to the punchline or *baufu*, which happens at the end of the fourth and final section (Tsau, 1979–1980, pp. 61–62). Moreover, XS always incorporates four elements into a performance: *shuo* (speaking), *xue* (mimicry), *chang* (singing), and *dou* (the provoking of laughter) (Lawson, 2011, p. 117). While the first three elements refer to the way performers manipulate their voices, the fourth implies the two ways in which actors may relate to each other to provoke an audience to laughter. In the most common style of XS known as "heavy on one end" (*yitouchen*), the main actor (*dougende*) plays the dominant role to the *penggende*, who plays the ostensibly subsidiary role as the straight man (p. 119). Nevertheless, even when he is perceived as a supporting actor, the *penggende* is recognized as wielding considerable power (Moser, 1990, p. 46), a phenomenon in which the power of the subordinate challenges the perceived dominance of the main actor (Lawson, 2017, pp. 91–95). In the second style of XS known as "two sides of a snap" (*zimugen*), the roles between the two actors are equal (Lawson, 2011, p. 120).

Given the way the Cambridge authors used ELAN to transcribe and analyze conversational bouts between same-sex friends, I wondered if the same software could be used to analyze a recording of XS to determine differences in rhythmicity (the regular use of rhythmic cycles) between the actors in the "heavy on one end" as opposed to the "two sides of a snap" styles. Additionally, because singing is one of XS's key elements, I was also interested in looking at the ways the two actors negotiated the differences between speech and song in performance. Since audience response is also openly acknowledged as foundational to comedic performance in XS (Liu, 1985; Xue, 1986), how exactly might the rhythmic punctuation of audience laughter factor into the musical and verbal volleys of the actors?

"A Carefree Life". Inspired by the Cambridge study and what I had learned about XS from the field, I began the process of looking for a recent performance of XS and located "*Xiaoao Jianghu*" (A Carefree Life) on youtube, featuring the popular actors Guo Degang and Yu Qian in a performance originally published on February 15, 2016. While I personally did not work with or interview either Guo or Yu, they are currently recognized as popular performers who have been particularly successful in updating XS to appeal to younger audiences of Millennials (Cai, 2016), so the recording provided a contemporary example for the project. The title of "A Carefree Life" is borrowed from the title of a serialized novel from a Hong Kong newspaper in the late 1960s (Lawson et al., 2020), and Guo's interpretation of "carefree" is reflected in his satirical treatment of elite Chinese cultural traditions, reflecting a casual and even dismissive attitude towards high culture. His hilarious portrayals of Chinese classical art forms lead to the *baofu* or climax of the performance in which he sings the concluding line of an especially difficult and notoriously high-pitched opera aria. However, his final triumph only comes after a painful process of playfully harassing Yu Qian, who acts as a foil to Guo throughout the performance. As Yu's most important task, he is supposed to set up Guo's final line of the aria, but appears to be genuinely afraid that his assignment—the first part of the line—is too high. Moreover, Guo teases Yu by baiting him. At first Guo exacerbates Yu's fears by singing too low and then in a raspy-sounding voice, both times eliciting strong laughter from the audience as they witness Yu's discomfort. After these false starts, Yu finally agrees to sing the first line, Guo triumphantly completes the aria, and the audience erupts into loud laughter and sustained cheering applause, marking the conclusion of the performance.

Case study: the XS project
The purpose of the XS study (Lawson et al., 2020) was to look at both pulse and pitch in the relationship between the two performers, which, at least superficially, appeared to mimic the relationship between two attitudinally aligned speakers in the successful conversational bouts highlighted in the Cambridge study. Additionally, we wanted to look at the relationships between the performers and the audience in both spoken and musical exchanges during the performance—something that was not relevant in the Cambridge study. We noticed that in some ways the XS performance seemed to fit Turino's definition of a presentational performance, in which one or more performers provide music (or, in this case, speech and music) to seated audience members who are not perceived to be in a performance role. After watching the recording of "A Carefree Life," however, the performance appeared to be presentational at times and participatory at other times. We also wondered how and if the distinctions between "heavy on one end" and "two sides of a snap" styles of interaction between the two performers might be manifested in the performance.

Even though we were looking to the Cambridge Study as a model, there were some major differences in performance and methodology between the two projects. First, although the scripted dialog between the XS actors is unlike the extemporaneous interactions between pairs of same-sex speakers in the Cambridge study, the audience reactions to the actors were both spontaneous and integral to the performance, constituting an element not present in the Cambridge study. Second, rhythmicity—the presence of regular rhythmic cycles in the spoken dialog—functioned differently in XS than in the interactions between speakers in the Cambridge study because of the presentational nature of a XS performance. In XS, audience response is constrained by the actors' performance and, therefore, does not demonstrate rhythmicity throughout the performance in the same way a successful bout demonstrates periodicity in the Cambridge study (see Clayton (2007) for a similar argument about audience response in a presentational performance setting). Moreover, since the meaning and execution of presentational speech is the focus of the performance in XS, the intelligibility of the actors' speech sometimes supersedes the rhythmicity that might have otherwise occurred in regular spoken interactions (see Hawkins (2014)).

Description of the XS study. In addition to using ELAN, we also used Praat acoustic analysis software and R statistical software to analyze the fluctuating relationships between presentational and participatory aspects of performance and between music and speech (Lawson et al., 2020). We were especially interested in the ways in which the audience became a third performing agent in a presentational setting and the tendency for pitch approximation in the final bout of the XS performance as a result of increased affiliative involvement between audience members and performers. The details of the study, including a complete translation of the text, a comprehensive description of the methodology and analysis of the data gleaned from ELAN and Praat, and a statistical analysis of the findings in R from both bouts are found in Lawson et al. (2020), so that information will not be duplicated here. Instead, I will summarize what we observed and measured during the two short parts of the performance—or "bouts," to use the same terminology from the Cambridge Study–we analyzed: Bout 1 occurred during the first part of the XS recording and Bout 2 occurred at the very end of the recording during the *baufu* or climax.

Bout 1. The first bout features only speaking with no singing (from minute 1:00 to minute 2:14), and showcases Guo's delusions about what his career would be like at age 140—a good example of the kind of hyperbole that is expected in a XS performance (Li S (1985) A brief study of the writing of traditional Xiang Sheng texts. Tianjin, China (unpublished paper)). The relationship between Guo Degang and Yu Qian during this bout is an example of the "heavy-on-one-end" style of XS (or a presentational style, to use Turino's terminology), meaning that Guo Degang is clearly dominant, with Yu Qian functioning as his foil. In the same way that the interlocutor responds with a type of backchannel-like commentary to the main speaker in a successful spoken bout in the Cambridge study on pairs of same-sex friends (Robledo et al., 2016), so this style of XS can only be successful with the proper supportive co-narration provided by the straight man (Moser, 1990). Audience laughter also becomes increasingly prominent by the middle of the bout, with the durations of audience responses even surpassing the durations of Yu's utterances. The climax of the bout occurs when Guo claims that his

acting partner, Yu, will still be performing with him at his advanced age. When Yu asks how he will still be there, Guo points to an urn on the stage—the one that will carry Yu's cremated remains. At this point there is significant audience response as they laugh at Guo's joke about Yu, signaling the end of Bout 1. The audience's responses increase in number and duration, eventually exceeding the length of Yu's responses to Guo. The audience is clearly a major participant in this bout, albeit different from the way Turino explains participatory interaction.

Although we did not find regular rhythmic cycles throughout the two bouts, we saw rhythmicity for approximately 50% of the bouts (Lawson et al., 2020). For the pitch data, we did notice an occasional matching of pitches (Lawson et al., 2020); however, when considering all of the pitch relationships between turns in the entire Bout 1, we did not identify pitch matching of any statistical significance.

Bout 2. The second bout is a battle of musical pitches, showcasing both Guo and Yu taking turns singing a difficult opera aria (from minute 12:50 to minute 14:57) at the end of the performance. In this bout the two performers have clearly switched to a two-sides-of-a-snap style of interacting in which both actors participate more equally than they did at the beginning of the performance—a situation which demonstrates a more participatory, rather than presentational, style of performing. Moreover, the audience also takes a more significant role in this bout, further contributing to the participatory nature of the bout. Witnessing Yu's fear of having to sing too high and Guo's relentless teasing, the audience becomes intimately involved in their drama during this second bout, wondering whether Yu's apprehension or Guo's brash confidence will be justified. Playing with the audience's emotions fuels the excitement, and when Guo finally sings the highly anticipated line, the audience erupts into wild cheering applause.

When analyzing the pitch and rhythmic data from ELAN, we considered each juncture between the pitch, measured in HZ, at the end of an utterance by one agent (Guo, Yu, or audience) and the initial pitch of the following response by the subsequent agent, similar to the turn transitions studied by the Cambridge group as outlined in Robledo et al. (2016). The data consisted of the ordered pairs of numbers in which the first number in the pair was the pitch of the last syllable of the first utterance by Guo, Yu, or the audience, and the second number was the starting pitch of the following response by Guo, Yu, or the audience. The analysis consisted of calculating the linear correlation coefficient between the ordered pair of numbers, and fitting a linear model where the dependent variable was the starting pitch (in Hz) of the following response. The independent variables were (1) the pitch of the last syllable of the first utterance, (2) the indicator that distinguished whether the audience or one of the actors gave the first utterance, and (3) the interaction between the indicator variable and the pitch of the last syllable in the first utterance.

In Bout 2 there was a highly significant linear correlation between the pitch of the last syllable of the first utterance and the pitch of the first syllable of the response (Lawson et al., 2020) in *both* the spoken and sung phrases. Thus, the anticipatory and improvisational interactions between actors and audience members—through reciprocal pitch approximations and long audience responses—energizes the second bout, demonstrating a participatory kind of interaction, rather than the presentational style at the beginning of Bout 1. In sum, Guo and Yu's performance gradually becomes more participatory as it builds to the punchline, all the while fueling increased participation on the part of the audience. Another way of looking at the success achieved in "A Carefree Life" is to see a gradual change from

Guo's initial presentational style to a more participatory style in which Guo, Yu, and the audience contribute more equally.

Referring back to Dissanayake's notions of signal and contribution, which I have renamed as presentation and participation, one could look at "A Carefree Life" in a couple of different ways. Borrowing Berliner's term, Guo and Yu's interaction could be seen as a presentation-participation loop in which the interaction between the two actors is the primary focus. However, as Yu's support allows Guo to gain momentum, the audience responds to the pair's evolving interactions through cheering applause, thereby also participating and contributing to Yu and Guo's performance as they interact with one another: Guo & Yu's presentation→audience participation→Guo & Yu's continued performance.

Thus, the findings from this preliminary study suggest that there is a measurable degree of rhythmicity for 50% of the performance and, even more strikingly, pitch approximation between audience and performers during the second bout, demonstrating a relationship between the performers' presentational signals and the audience's participatory response by the climax of the performance—findings that could be tested by other researchers studying similar kinds of recorded performances.

"Attitudinal alignment" as reception. In the Cambridge study, the high level of emotional engagement (referred to as attitudinal in the Cambridge Study) that occurs between same-sex friends appears to be a prerequisite for what the researchers call a successful bout (Hawkins et al., 2013; Robledo et al., 2016; Cross et al., 2016). Similarly, it appears that a successful XS performance must demonstrate a strong affiliative relationship—an example of attitudinal alignment—between the audience and the performers, as well as between the performers themselves. However, not all XS performances are as well received as the 2016 youtube version of "A Carefree Life." In the field, I attended a few XS performances in which the audience did not applaud at all—a rather brutal response, to be sure, but one that reflects the belief that performers must woo their audience and meet their expectations in order to merit applause (Lawson, 2011, pp. 18–20). In other words, in the cases where XS performers do not emotionally engage—or attitudinally align—with their audience, there would be little or no connection with the audience and, therefore, no rhythmicity or pitch approximation between actors and audience members as seen in the highly engaged concluding section of "A Carefree Life."

The challenge is how to ascertain, measure, and analyze the ways in which audience members *receive* a performance before they respond. In any given performance there will be those who are unengaged, some intensely engaged, and others who fall somewhere between the extremes of that spectrum. For example, Jonathan Stock (2017) makes the following comments about the different members of an audience for a Chinese opera:

> The audience beside you, then, is hardly one unified listenership, except insofar as you've all chosen to attend this performance and gained entrance to it. Factors like personal outlook, expertize, dedication, dress and gender all play significant roles in shaping how you're listening to the opera performance and how you view the other people who surround you. You can readily imagine that there must be a correspondingly varied set of expectations among the performers, and among those others whose input further determines the framing and character of the performance event, from promoters to censors, and from directors to costume designers… (2017, p. xiii).

Thus, the differing backgrounds and expectations of the audience members contribute to a set of varied responses to the

opera being performed on stage—a scenario that contrasts greatly with the seemingly unified responses of the audience members on the recording of "A Carefree Life."

The question about the level of engagement hearkens back to the second element raised by Dissanayake: the reception of the signal. Yet the idea of "receiving" a performance is especially difficult to measure, particularly during a performance. In the case of the XS study, the audience appeared to respond as a single entity to the performance, allowing us to measure the auditory response generated by their collective cheering applause, recorded by Praat and ELAN as a unified response. Aside from the fact that there were enough people responding en masse to produce a clear audio signal, there may have been some in the audience who did not respond with the rest of the crowd. How, then, might researchers determine the different responses of audience members, particularly those who did not cheer and applaud?

In addition to interviewing listeners after the fact[2], another way to discover the various ways audiences respond to performance is in real time through interactive technologies found in facilities such as LiveLab at McMaster University—a performance space that has the capacity to measure the responses of audience members in terms of movement (motion capture), brain responses (EEG), muscle tension (EMG), heart rate, breathing rate, and sweating (GSR) for select members of the audience in seats with equipment to take those measurements[3]. However, the equipment is expensive, invasive, not transportable for use in the field, and, therefore, not practical for use in a typical concert or in a field setting. Nevertheless, the ability to measure the initial, pre-applause responses of a select number of audience members—even under such contrived circumstances—allows us to understand something more about the different physiological reactions involved in the process of engaging with a performance *before* hearing or observing any outward behavior—an aspect of performative mutuality that has not yet been adequately understood or studied by ethnomusicologists. Why might this be important?

Some scholars have remarked that Western European classical music is a presentational kind of music whose audiences are quiet and passive, compared to the participatory nature of the music produced in cultural settings in which audience members are more actively engaged in the performance (Becker, 2004, p. 69; Turino, 2008, p. 26; Tsioulakis and Hytonen-Ng, 2017, p. 3). However, the claim about audience passivity in response to Western classical music is based on a superficial and limited examination of outward audience reactions. As seen in the documentary on Joe Heaney, periods of silent listening do not necessarily mean that the audience is passive or unengaged. On the contrary, the *sean-nos* clip indicates a deep level of absorption on the part of the listeners. Further, the XS study demonstrates that a performance can, at times, involve a seated audience that is silent during stretches of presentation; however, instances of seemingly silent reception may be followed by loud, exuberant audience participation, indicating that audiences may still be actively engaged even if they are not constantly responding visibly or audibly to the performers. Thus, emotionally charged applause following a *sean-nos* performance or a performance of Western classical music demonstrates how an apparently "passive" audience might be highly engaged, even if not always vocally responsive, throughout the performance; or, in the case of XS, audiences might cheer and applaud *intermittently* during a performance. In the case of a poorly executed performance and/or an unengaged audience, there may appear to be no response whatsoever. In that case, the audience might be unresponsive and unmoved or frustrated at not having their expectations met.

Ascertaining and measuring the different ways in which audience members receive performances, then, is an untapped area for future research with promise for gaining a more nuanced understanding of different types and levels of performative mutuality. Conceivably, wearable technologies will become more advanced and widespread, supplanting the expensive, cumbersome and invasive equipment that is currently being used to measure audience reception at present.

Conclusions

Cross's claim that music and language are two halves of the human communicative toolkit (2009, 2014) is well supported by Chinese XS, and it appears that when semantic information is paramount, as in much of the XS performance, the spoken mode dominates (Hawkins, 2014). However, XS performances also contain strong musical elements because "*chang*," or the singing component that is part and parcel of crosstalk, is a signature part of each performance. It is no wonder that the traditional rubric under which XS falls is known as *shuochang* or the "speaking singing genres." The obvious way song is used intermittently with spoken modes in XS is not only essential to the performance; however, the discovery of a hidden musicality in the banter between performers and audience members offers additional evidence of the reciprocal nature of speech and song in human communication and supports Patel's claim that a study of the relationship between music and language is relevant to scholars from both scientific and humanistic backgrounds.

Guided by Dissanayake's three points, analysis of a comedic performance in which music and language are both essential elements yielded a fascinating result: the reciprocity between the presentational signals of performers and the contributory responses of the audience was most pronounced at the end of the sung portion, which was the most emotionally charged point of the performance. Might a strong emotional connection involve increased "musicality" in terms of pitch alignment and periodicity? Results from this preliminary project tentatively suggest that when interactants are attitudinally aligned—and therefore emotionally engaged—in a XS performance, speech may become (1) rhythmically entrained for about 50% of the performance and (2) melodically entrained in terms of pitch approximation through the process of co-narration by the end of the performance. Much more research is needed in order to begin to generalize about rhythmic entrainment and pitch matching across turns in comedic performance, but the analysis of a videotaped recording using ELAN and Praat is something that could be replicated in studying other recordings of comedic performance traditions in which there are strong, measurable audio signals from the audience.

As a corollary to the previous point, this study also challenges the way scholars consider presentational vs participatory interactions in musical performance (Cross et al., 2016; Turino, 2008). In the recording of XS, the performed speech of the actors at the beginning of the performance eventually changed from a presentational to a participatory mode as Yu and the audience members became gradually more involved. Based on the changing modes of interaction in this crosstalk, presentational and participatory interactions are not mutually exclusive, but rather complementary and fluid, reflecting a gradual change from what the Chinese call the "heavy-on-one-end" mode of interaction, which corresponds to a presentational style, to the "two-sides-of-a-snap" performance that is similar to a more participatory style of performance.

Finally, Dissanayake's second point about the way a signal is received is one of the areas in which wearable technologies and an empirical approach to musical performance could add considerable depth to our observations about audience participation. By ascertaining, documenting, and measuring the different ways

audience members receive incoming presentational signals, researchers can begin to understand how some audience members respond enthusiastically throughout a performance, others respond intermittently, and still others only at the end—or not at all. Responses to a presentational signal vary dramatically and the nature of those responses should figure into the way we study performative mutuality in music.

In conclusion, this paper responds to the call issued by the participants in the 2018 Frankfurt conference (1) to expand the purview of music cognition research by including music and participants from cultures outside of the Western European musical tradition and (2) to explore ways scholars from different disciplinary backgrounds can be encouraged to consider tools and methodologies used by other disciplines and increase collaborative research opportunities. First, the paper presents an example from a culture and a genre that is outside of the Western European musical tradition, providing additional intercultural perspectives in the study of music cognition. Second, the paper offers a case study as to how a particular research problem—the interface between music and language in Chinese musical comedy —may benefit from utilizing a variety of unexpected disciplinary perspectives. Although Jacoby et al. (2020, p. 188) imply that there is an impasse between experimental rigor and ecological validity, this paper suggests that empirical and ethnomusicological paradigms might work best when they are not used simultaneously, but rather in tandem: the XS Study was rooted in an ethnomusicological inquiry that benefitted from considering the results of and tools used in a previously conducted empirical study. Ideally, each methodological approach could provide results and questions for the other to pursue, provoking researchers from both disciplines to refine their respective approaches in subsequent studies. If the premise behind performative mutuality were to extend to scholarship, we would be able to create a scholarly mutuality between disciplines that would stimulate more research than if either discipline only functioned independently.

Notes

1 An exception is Wennerstrom's plea for considering prosody in discourse analysis, claiming that prosody demonstrates how we are "creative musicians in the symphony of communication that forms the basis of our lives as social beings" (2001, p. 263).
2 See Gabrielsson and Lindstrom Wik (2003) and Herbert (2011) as two examples of interviewing peoples' differing responses to music.
3 For more information about this facility, go to https://science.mcmaster.ca/tour/psychology-neuroscience-behavior/location-live-lab.html.

References

Allport G, Vernon P (1933) Studies in expressive movement. Macmillan Company, New York, NY
Ambady N (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. Psychol Bull 111(2):256–274
Becker H (1951) The professional dance musician and his audience. Am J Sociol 57 (2):136–144
Becker J (2004) Deep listeners: music, emotion and trancing. Indiana University Press, Bloomington
Becker J (2009) Ethnomusicology and empiricism in the twenty-first century. Ethnomusicology 53(3):478–501
Berliner P (1994) Thinking in jazz: the infinite art of improvisation. Chicago studies in ethnomusicology series. University of Chicago Press, Chicago
Cai S (2016) A culture hero: Xiangsheng (Crosstalk) performer Guo Degang. Asian Theatre J 33(1):82–103
Clayton M (2007) Time, gesture and attention in a Khyal performance. Asian Music 38(2):71–96
Cross I (2009) Music as a communicative medium. In: Botha R, Knight C (eds) The prehistory of language. Oxford University Press, Oxford, pp. 113–134
Cross I (2014) Music and communication in music psychology. Psychol Music 42 (6):809–819
Cross I, Ogden R (2013) Communicative interaction in spontaneous music and speech. In: Orwin M, Howes C, Kempson R (eds) Music, language, and interaction. college. Publications, London, pp. 285–329
Cross I, Fitch WT, Aboitiz F, Iriki A, Jarvis ED, Lewis J et al. (2013) Culture and evolution. In: Arbib M (Ed) Language, music and the brain. Strungmann forum reports. MIT Press, Cambridge, pp. 541–562
Cross I, Hawkins S, Ogden R, Robledo J P, Gill S, London J, Ross B, Liu J, Knight S, Halford D (2016) Music as communication, communication as music. Paper presented at the annual meeting of the Society for Ethnomusicology, November, 2016, Washington, D.C
Davitt M (1996) Joe Heaney: sing the dark away. TV Movie Documentary
Dissanayake E (2010) Root, leaf, blossom, or bole: concerning the origin and adaptive function of music. In: Malloch S, Trevarthen C (eds) Communicative musicality: exploring the basis of human companionship. Oxford University Press, Oxford, pp. 17–30
Fatone G, Clayton M, Leante L, Rahaim M (2011) Imagery, melody and gesture in cross-cultural perspective. In: Gritten A, King E (eds) New perspectives on music and gesture. Ashgate, Farnham, England, pp. 203–220
Gabrielsson A, Lindstrom Wik S (2003) Strong Experiences related to music: A Descriptive System Musicae Scientiae. 7(2):157–217
Frith S (1998) Performing rites: evaluating popular music. Oxford University Press, Oxford
Goffman E (1979) Frame analysis: an essay on the organization of experience. Harvard University Press, Cambridge
Guo D and Yu Q (2016) A Carefree Life https://www.youtube.com/watch?v=kobiKbm5g_o
Hamilton A (2007) Lee Konitz: conversations on the improviser's art. University of Michigan Press, Ann Arbor
Hawkins S (2014) Situational influences on rhythmicity in speech, music, and their interaction. Philos Trans R Soc Lond B: Biol Sci 369(1658):20130398
Hawkins S, Cross I, Ogden R (2013) Communicative interaction in spontaneous music and speech. In: Orwin M, Howes C, Kempson R (eds) Music, language, and interaction. College Publications, London, pp. 285–329
Herbert R (2011) Everyday music listening: absorption, dissociation and trancing. Ashgate, Burlington, Vt
Hersch C (2019) Unfinalizable: dialog and self-expression in jazz. In: Gebhardt N, Rustin-Pashal N, Whyton T (eds) Routledge companion to jazz studies. Routledge, New York, NY, pp. 367–376
Hesmondalgh D, Negus K (eds) (2002) Popular music studies. Arnold, London
Honing H (2018) The origins of Musicality: interdisciplinary perspectives on the capacity to perceive, appreciate, and make music. The MIT Press, Cambridge
Hytonen-Ng E (2017) Contemporary British jazz musicians' relationship with the audience: renditions of we-relations and intersubjectivity. In: Tsioulakis I, Hytonen-Ng E (eds) Musicians and their audiences: performance, speech and mediation. Routledge, London, pp. 69–85
Jacoby N, Margulis E, Clayton M, Hannon E, Honing H, Iverson J, Klein T, Mehr S, Pearson L, Peretz I, Perlman M, Polak R, Ravignani A, Savage P, Steingo G, Stevens C, Tainor L, Trehub S, Veal M, Wald-Fuhrmann M (2020) Cross-cultural work in music cognition: challenges, insights, and recommendations. Music Percep 37(3):185–195
Johnson B (2017) In the body of the audience. In: Tsioulakis I, Hytonen-Ng E (eds) Musicians and their audiences: performance, speech and mediation. Routledge, London, pp. 15–33
Lawson F (2011) The narrative arts of tianjin: between music and language. Ashgate Publishing, Farnham
Lawson F (2017) The women of Quyi: liminal voices and androgynous bodies. Routledge, New York, NY
Lawson F, Sims J, Lawson J (2020) When audiences become performers and speech becomes music: new tools to analyze speech, song, and participation in Chinese crosstalk. Music Sci 3:1–18
Leante L (2017) Observing musician/audience interaction in North Indian classical music performance. In: Tsioulakis I, Hytonen-Ng E (eds) Musicians and their audiences: performance, speech and mediation. Routledge, London, pp. 34–49
Link P (1980) Hou Baolin: an appreciation. Chin Lit 2:84–94
Liu Z (1985) Tan Xiangsheng Jiezou (a discussion of the rhythm in Xiangsheng). Tianjinshi wenhuaju xijuyanjiushi, Tianjin
Malloch S, Trevarthen C (2010) Musicality: communicating the vitality and interests of life. In: Malloch S, Trevarthen C (eds) Communicative musicality: exploring the basis of human companionship. Oxford University Press, Oxford, pp. 1–11
Malotki E, Dissanayake E (2018) Early rock art of the American west: the geometric enigma. University of Washington Press, Seattle

Margulis EH, Wong PCM, Simchy-Gross R, Mcauley D (2019) What the music said: narrative listening across cultures. Palgrave Commun 5:146

Mehr S, Singh M, Knox D, Ketter DM, Pickens-Jones D, Atwood S et al. (2019) Universality and diversity in human song. Science 366:eaax0868

Middleton R (1990) Studying popular music. Open University Press, Milton Keynes

Moser D (1990) Reflexivity in the Humor of *Xiangsheng*. CHINOPERL Papers 15:45–68

Nettl B (2005) The study of ethnomusicology: thirty-one issues and concepts, 2nd edn. University of Illinois Press, Urbana

Patel A (2008) Music, language, and the brain. Oxford University Press, Oxford

Robledo JP, Hawkins S, Cross I, Ogden R (2016) Pitch-interval analysis of 'periodic' and 'aperiodic' Question+Answer pairs. In: Barnes H, Brugos A, Shattuck-Hufnagel S, Veilleux N (eds) Proc Speech Prosody 8: 1071–1075. http://www.isca-speech.org/archive/SpeechProsody_2016/

Savage PE (2019) Cultural evolution of music. Palgrave Commun 5:16. https://doi.org/10.1057/s41599-019-0221-1

Schlegloff E (2007) Sequence organization in interaction: a primer in conversation analysis, Vol. 1. Cambridge University Press, Cambridge

Sidnell J (2010) The ordinary ethics of everyday talk. In: Lambek M (ed) Ordinary ethnis: anthropology, language, and action. Fordham University Press, New York, NY, pp. 123–139

Sidnell J, Stivers T (eds) (2012) The handbook of conversation analysis. Wiley-Blackwell, West Sussex

Small C (1998) Musicking: the meanings of performing and listening. Wesleyan University Press, Middletown, CT

Stock J (2017) Audiencing. In: Tsioulakis I, Hytonen-Ng E (eds) Musicians and their audiences: performance, speech and mediation. Routledge, London, pp. xii–xiv

Tolbert E (2012) An ethnomusicological perspective on animal "music" and human music: the paradox of the "paradox of rhythm." In: Reuschat P, Rohrmeier M, Hawkins J, Cross I (eds) Language and music as cognitive systems. Oxford University Press, Oxford, pp. 121–127

Trevarthen C, Malloch S (2000) The dance of wellbeing: defining the musical therapeutic effect. Nordisk Tidsskrift Musikkterapi 9(2):3–17

Tsau S (1979–1980) Xiangsheng and its star performer, Hou Baolin. CHINOPERL Papers 9:32–79

Tsioulakis I, Hytonen-Ng E (2017) Introduction. In: Tsioulakis I, Hytonen-Ng E (eds) Musicians and their audiences: performance, speech and mediation. Routledge, London, pp. 1–12

Turino T (2008) Music as social life: the politics of participation. University of Chicago Press, Chicago

Turner E (1977) Communitas: the anthropology of collective joy. Kamgrave Macmillan, New York, NY

Wennerstrom A (2001) The music of everyday speech. Oxford University Press, Oxford

Widdess R, Finkensiep C, Rohrmeier M (2019) Modelling the syntax of north indian melodies with a generalized graph grammar. Proceedings of the International Society for Music Information Retrieval Conference. Delft, Netherlands, pp. 462–469

Xue B (1986) Interview by author. Nankai University, Tianjin, China

Acknowledgements

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.R.S.L.

Reprints and permission information is available at http://www.nature.com/reprints

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