Evaluating the Audience's Perception of Real-time Gestural Control and Mapping Mechanisms in Electroacoustic Vocal Performance

J. Cecilia Wu
MAT, UC Santa Barbara. CA
93106, U.S.A.
jcw@umail.ucsb.edu

Madeline Huberth,
Yoo Hsiu Yeh
CCRMA, Stanford University, CA
94305, U.S.A.
mhuberth@ccrma.stanford.edu
yoohsiu@alumni.stanford.edu

Matthew Wright
CCRMA, Stanford University, CA
94305, U.S.A.
matt@ccrma.stanford.edu

ABSTRACT

This paper presents an empirical evaluation of a digital music instrument (DMI) for electroacoustic vocal performance, the Tibetan Singing Prayer Wheel (TSPW). Specifically, we study audience preference for the way it maps horizontal spinning gestures to vocal processing parameters. We filmed six songs with the singer using the TSPW, and created two alternative soundtracks for each song: one desynchronized, and one with the mapping inverted. Participants viewed all six songs with either the original or desynchronized soundtrack (Experiment 1), or either the original or inverted-mapping soundtrack (Experiment 2). Participants were asked several questions via questionnaire after each song. Overall, they reported higher engagement and preference for the original versions, suggesting that audiences of the TSPW prefer more highly synchronized performances, as well as more intuitive mappings, though level of perceived expression of the performer only significantly differed in Experiment 1. Further, we believe that our experimental methods contribute to how DMIs can be evaluated from the audience's (a recently noted under-represented stakeholder) perspective.

Author Keywords

Tibetan Singing Prayer Wheel, NIME evaluation, audience's perception, electroacoustic vocal performance, gesture mapping,

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing — Methodologies and techniques; H.5.2 Information Interfaces and Presentation (e.g., HCI): User Interfaces - Evaluation / methodology.

1. INTRODUCTION

The Tibetan Singing Prayer Wheel (TSPW) is a handheld, wireless, sensor-based musical instrument that simultaneously processes vocals and synthesizes sound based on the performer's hand gestures with a one-to-many mapping strategy, introduced in NIME 2015 [31].

We designed the TSPW to let electroacoustic vocalists augment their vocal performance in real time, using synchronized hand gestures, an intuitive mapping strategy, and a solo instrument to achieve performance goals that would normally require multiple

instruments and activities. Our previous work evaluated the TSPW's latency, reliability, reproducibility, and hardware optimization. In this work, we design and test a methodology for evaluating TSPW's perceived musical expressiveness from the audience's perspective, with the primary goal of validating the mapping design and control strategies for the TSPW's real-time voice processing. We frame this study within the context of evaluation of DMIs and performers gesture as it relates to sound, and make efforts to be clear about our stakeholders, evaluation goals, criteria, methodology, and duration, as suggested in a recent survey of NIME papers [2].

Figure 1. A Tibetan Singing Prayer Wheel interface

2. AUDIENCE PERSPECTIVE OF DMIs

Typical DMI evaluation case studies [14][17][26] primarily examine an instrument's usability and perceived musical expression from the performer's/player's perspective, using or based on the framework of Wanderley and Orio [29], which adapts human-computer-interaction user-study methodology to musical tasks. While the performer's perspective is a critical one, other stakeholders such as the audience, designer, and manufacturer are also important. In 2011, O'Modhrain proposed a framework for DMI evaluation from multiple evaluation perspectives, goals, and stakeholders [20], enumerating several research studies that focused on a perceivable causal link between the gestural input and control mechanism and its produced sound [7][25]. She pointed out that it is specifically important to study the audience's perspective in order to evaluate the gesture-sound relationship.

Based on O'Modhrain's framework, in 2012, Barbosa et al. proposed a meaningful methodology in a case study evaluating the audience's perspective of a DMI named "Illusio." In 2013, Barbosa et al. further evaluated the same DMI with the stakeholders of both players and audience [1]. Jordà and Mealla proposed "a conceptual

framework that could serve in evaluating the potential, the possibilities, and the diversity of new digital musical instruments, focusing on the expressive possibilities these instruments can offer to their performers." They specifically looked into the mapping strategies and expressiveness relationship, from the audience/listener perspective [15]. Our evaluation of the TSPW also focuses on the audience's perspective, since for live vocal performances and music making, the "measure of success is the response of the audience to their performance" [20].

2.1 Evaluating Gesture and Vocal Processing

Many new musical interfaces enhance human vocal expression using gestural control as well as diverse design, input, mapping, and control strategies [4][13][19]. However, to the best of our knowledge, no follow-up scientific evaluation has been conducted. Musical gestures in computer music composition and DMI design have been well studied since the 1980's [5][9][12], but not in the specific context of electroacoustic vocal performance.

Vocal performance is unique from other instrument performance in several ways. First, sound comes directly from a vocalist's body; there is no other sound generator. Moreover, humans can naturally read body language and voices, beginning at infancy and refined to an art by adulthood [6]. Hence the audience perceives more complex cognitive meanings (communicating emotions and expressing musicality) from a vocalist's body movement and gestures as compared to other musicians' [16].

The TSPW affords two main gestures: vertical (raising and lowering the device) and horizontal (speed of the wheel's spinning motion). Vertical height maps directly to the pitch of the synthesized sound. Generally, previous empirical work shows a strong perceptual correspondence between ascending pitch and ascending height [21][22][18] , across a broad population, "applying to both musicians' and non-musicians' conscious and unconscious cognitive processes" [10]. Especially in [18], musicians and non-musicians were asked to move their arm in response to highly controlled stimuli varying in pitch, loudness and tempo. Their results reported higher pitch leading to higher elevation in space and increasing tempo. Thus, we focused our evaluation on the less conceptually and empirically explored gesture: the horizontal spinning motion. This gesture is uncommon among DMIs, and we are unaware of studies evaluating the kinds of processing that might be associated with it. Thus, one of our primary evaluation goals was to validate the TSPW's designed mapping from horizontal spinning to real-time vocal processing.

3. RESEARCH QUESTIONS

3.1 Synchronicity of Processing

Within the NIME field, where the researchers build instruments, it is generally taken as obvious that the performer's' gestures with these instruments should (directly or indirectly) determine the sonic output behavior of these instruments, and also that the effect of the gesture should track the gesture with the lowest possible latency and jitter. We call this the "synchronicity" of the mapping, and are not aware of any prior work that questions or validates the assumption that synchronous NIME mappings are actually better from the audience's perspective. Our two conditions for examining this assumption, detailed in section 4.2.1 are: 1) processing tracks the gestures versus 2) gestures are non-relevant to the processing.

3.2 Intuitiveness of Mapping

Generally in instrumental performance contexts, faster motions produce faster and/or more energetic sound.

Listeners also, given the sound alone, have associations with how *imagined* movement relates to the sound, and it has been shown that a change in attack rate is positively associated with change of speed of imagined movement [10]. Therefore, we infer that faster horizontal spinning hand motions would be expected to produce more intensely processed sounds (in our case, more intensely granulated), and vice versa. This is the TSPW's original mapping strategy, and we refer to it as "intuitive gesture mapping".

3.3 Hypotheses

Our Experiment 1 ("synchronization evaluation") and Experiment 2 ("intuitive mapping evaluation") test the following hypotheses:

1. Greater synchronicity in how gestures and hand/arm movements control vocal processing enhances electroacoustic vocal performance, resulting in increased perceived expression of the performance and increased audience engagement.

2. When arm/hand movements are intuitively matched with the vocal processing using the prayer wheel's horizontal spinning motion and speed variables as the control mechanism, the level of perceived expression of the performance and the engagement of the audience will be higher than when movements and processing are not intuitively matched.

3.4 Methodological Approach

The general methodology in our study is informed by studies on the relationships between performer expression and gesture from the perspective of the audience, for example [11][27]. Some of these studies' experimental design involves showing multiple versions of performances to audience members, who then provide judgments about the performances [8][11][3].

Some cases use videos of performers in which the same video of the performer is shown with audio manipulated, or the same audio applied to multiple videos, to see the effect of gesture on auditory perception. For instance, Vuoskoski et al. [28] investigated the relative contributions of auditory and visual kinematic cues in the perceived expressivity of piano performances by presenting matched and mismatched audiovisual information from 'normal,' 'exaggerated,' and 'immobile' performances. While in their study, videos were post-processed to align motions of different performances with various audio recordings, our study compares different processing versions of the same original raw vocal signal, so the fundamental timing of the performer's motions (her singing resulting in time-synchronized sound) line up with the video.

In both experiments, half of the videos contained our designed, original, synchronized vocal processing, where slow spinning mapped to long, reverberant vocal effects and fast spinning mapped to short, granular vocal effects. In Experiment 1 the alternate videos' vocal processing was desynchronized from the motion. In Experiment 2, the alternate videos contained an 'inverted' style of processing, in which faster spinning mapped to the long reverb vocal effect.

4. METHODS

Since the two experiments were similar in methodology and design, varying primarily by stimulus type, we present them side-by-side in this and the following Results section, with separate subsections for the stimulus design for each experiment.

4.1 Participants

In the synchronization-evaluation experiment, twenty-five musicians (9 female) and one non-musician (female) participated (mean age = 24.9 years, SD = 5.4 years). 'Musician' was defined as having 5+ years of formal musical training (following Vines et al., who used this training cutoff "to ensure a musical ear" [3]). The intuitiveness-evaluation experiment had twenty-two participants (10 non-musicians, 5 female, plus 12 musicians, 3 female; mean age = 29.3 years, SD = 9.1 years). Across both instantiations, the

vast majority of participants reported to be unfamiliar with Tibetan/Nepalese music at the time of taking this study: only four reported moderate familiarity. Musician participants had a wide variety of musical experience (e.g., vocals, piano, guitar, flute, drum set, cello, trumpet, oud, harp, gamelan, and drums). All participants were entered into a lottery for one of 5 $20 gift cards.

4.2 Musical Stimuli

Six vocal pieces were adapted for TSPW performance: 3 composed by the first author, 1 spoken word piece, and 2 Himalayan folk songs (M = 1.34 min, SD = 30 sec), representing a variety of styles and affects (rhythmic, melodic). We videotaped and audio recorded a performance of each piece by a singer fluent with the instrument (the first author) to eliminate potential variability in perceived expressiveness of different performers. We recorded the dry voice audio and the changing values of the parameters controlled by the horizontal spinning gesture as time-synchronized wav files.

The TSPW's vocal processing is a delayed and windowed granular effect with two parameters: pulse width (originally inversely related to speed) and pulse interval (originally proportional to speed) [31]. For each experiment two versions of the processing were made using a modified PureData patch: the original, and an alternative designed to test the study's hypotheses.

4.2.1 Synchronization Evaluation Stimuli

For the *"desynchronized"* version, the gestures and the vocal processing have no causal relationship and are hence not closely aligned. Every performance produced a unique pattern of TSPW rotational speed over time; each desynchronized soundtrack reprocessed the dry voice recording using the recorded control signals *from a different performance*. For example, the dry voice of song 1 was reprocessed with parameter envelopes taken from the performance of song 2, the dry voice of song 2 was reprocessed with parameter envelopes from song 3, etc. This effectively applied a final sequence that was uncorrelated with the original sequence, but with the applied time-varying control functions qualitatively similar to the "correct" control functions from the original performance (having been generated with the same instrument and by the same performer in similar performance contexts). For example, Figure 2 shows the pulse width control parameter as a function of time for compositions 2 and 3.
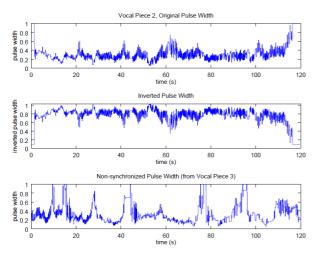

Figure 2. Top: Original TSPW-controlled pulse-width values used to process dry voice of piece #2. Middle: Inverted pulse-width values of piece #2. Bottom: Original TSPW-controlled pulse-width values of piece #3.

4.2.2 Mapping Evaluation Stimuli

For the *"inverted"* version, we inverted the recorded control signal as shown in Figure 2. The pulse width and interval parameters first were normalized, then subtracted from 1, effectively making previously small values large, and vice versa. In this inverted version, faster spinning hand gestures map to a longer pulse and thus a slower and longer reverb.

Creating the inverted version brought out the fact that the TSPW does not operate linearly at the ends of its range. Inertia and static friction make it physically impossible to spin the wheel extremely slowly, so on a linear scale from 0 (no spinning) to 1.0 (capped upper limit for performer's fast spinning, to maintain adequate dynamic range during standard use), the average minimum nonzero speed (over multiple trials) is 0.061. When inverting, this x=0 or x>0.06 behavior becomes x=1 or x<0.94. What was previously a jump between stopped and spinning now became a gap in which the parameters never smoothly approach the ceiling but rather clip at the x <0.94 maximum until jumping discontinuously to x=1. In addition, since the performer predominantly spun the TSPW at faster speeds with only a few complete stops throughout all 6 pieces, the inverted parameters rarely hit the maximum and contain longer periods at zero value. The initial inverted-parameter versions thus had overall more reverb than the original versions, so to make the inverted versions qualitatively the same as the original instrument we added an offset of 0.09 (determined empirically). Figure 2 compares the original and inverted pulse-width parameter for vocal piece #2.

4.2.3 Stimuli Presentation

For each song in each evaluation, there was an original video and an altered video, both using the same visual component and the same performance, but with a different audio track. For experiment 1, the altered soundtrack used the desynchronized parameters, and for experiment 2, the altered soundtrack used the 'inverted' parameters. Each participant randomly heard either the original or altered soundtrack for each piece, never viewing both versions of any piece. Thus, in the aggregate, there was an approximately equal number of viewings of each version of each performance, allowing participants to comment on differences in processing, if they detected any, while avoiding the potentially confusing situation of the same subject seeing both versions of the same performance.

4.3 Procedure

The experiment was administered via a custom Processing program, and was presented in a quiet room on a 15-inch Apple laptop using high-quality SONY MDR-NC8 headphones. At the beginning of the experiment, participants were told that there were different vocal processing methods being implemented with the evaluated instrument during various performances. Instead of focusing on the performer's ability, participants were asked to rate the performances and to pick their favorite(s) based on the vocal processing method and how the instrument was used. Each participant watched a total of 6 performances and answered 6 questions after watching each video. The questions were 5-point Likert-scale questions with the descriptive "extremely," "very," "moderately," "slightly," or "not at all," and were presented without a number associated. The questions were: "How expressive was the performer?", "How engaged were you in the performance?", "How effectively did the performer use the instrument to control the voice processing?", "How convincingly was the instrument responding to the performer's vocal expression?", "How closely aligned was the movement of the instrument and the voice processing?", and "How much do you like the way the instrument was changing the vocals?".

After all videos were presented, a demographic questionnaire was administered, and a final question was asked: "Please recall your favorite performance. What did you like about it? Do you think the instrument did or did not change the performer's ability to be expressive? Please elaborate if possible. If you recall which

video it was in order of presentation (e.g., first, second…), please state so." One experimental session lasted between 20-35 minutes.

5. RESULTS
5.1 Likert-Scale Questions
Since the Likert-scale questions resulted in ordinal (not continuous) data, we chose to analyze the responses from those who saw the original or altered versions of the songs using Fisher's exact test, which tests if there was a difference in distributions using contingency tables (i.e., does the distribution of responses to the questions differ depending on seeing an original or altered video?). This test is comparable to the chi-squared test, but was chosen instead since the sample size was small enough that the chi-squared test's dependence on a chi-square distribution would not hold.

For all questions in the synchronicity evaluation, the responses to the original version were significantly different to those of the altered (all $ps < .001$). For each question, those who saw the original version found it more frequently to be more expressive, frequently were more engaged, and both preferred and found more convincing the interaction between performer and instrument. In the mapping intuitiveness evaluation, the responses to Q2 - Q6 of the original version were significantly different to those of the altered (all $ps < .05$), while the response to Q1 was nearly so (p = .07). Figure 3 shows the count of responses based on movie version for each condition of the two evaluations, from 1 = "not at all" to 5 = "extremely". Group means are presented as dashed vertical lines as an additional visualization of the average response, since, even though these data are ordinal, the response labels have been shown to be close to semantically equidistant from each other [23]. The red band above each "5" indicates that subjects viewing the original version of the mapping were relatively consistently more likely to answer "extremely."

5.2 Free-Response Questions
The study's final question asked participants to reflect in writing on which performance was their favorite and why. We categorized their responses into 5 general types, presented in Table 1, along with their frequency in response to the original and altered videos. Some participants noted several videos in their responses, as well as more than one reason for selecting a favorite - we have included each reason discussed for each video mentioned in the table.

Notably, in the synchronization evaluation, subjects often justified their choice of a favorite based on the mapping's intuitiveness, never describing a desynchronized version as more intuitive (7/0). Interestingly, participants who justified their favorite performance in terms of the performer's particular expressiveness (without mentioning the instrument or its mappings at all) chose an altered video as often as not (4/4).

In the mapping evaluation, the most frequent reason to prefer a

video was that the gesture and processing were closely aligned, which matches with the experimental manipulation of these videos. In comparison to Experiment 1, this emerged as a more frequently reported factor in a video being a favorite, rather than the intuitive nature of the matching of the gestures and processing. Also notable is how few respondents mentioned how the instrument did not distract from the performance.

Table 1. Counts per movie type of categories of free-response answers to why a particular performance was preferred by the participant. O = Original, A = Altered; Expt 1 = Synchronization, Expt 2 = Mapping.

Response type	Expt 1		Expt 2	
	O	A	O	A
Instrument added to performance, instead of distracting from it	8	2	1	2
Very expressive by performer (no mention of instrument)	4	4	2	1
The song itself is beautiful and the processing matches the song	3	3	4	3
Think our mapping and processed vocal are more intuitive, thus making the piece most expressive than other mapping	7	0	0	0
The physical/visible intensity of the gesture is aligned (synchronized) with intensity in the effects processing	5	0	9	0

6. DISCUSSION
6.1 Synchronization Evaluation
The NIME community's assumption about using synchronous control and mapping strategies in designing DMI seems reasonable. The tight synchronicity between movement and sound is implicit in the design of NIME-style instruments [30]. Unsurprisingly, audience members tended not to favor our straw man non-synchronous mapping where the movement does not affect the sound processing. Indeed, non-synchronicity seemed to have widespread effects on participants' responses. When watching the original videos, participants found themselves more engaged, the performer more expressive, and liked the processed voice and performances more than when they watched the altered videos. When using the desynchronized processing and mapping, some audience perceived the prayer wheel as processing the vocals in a
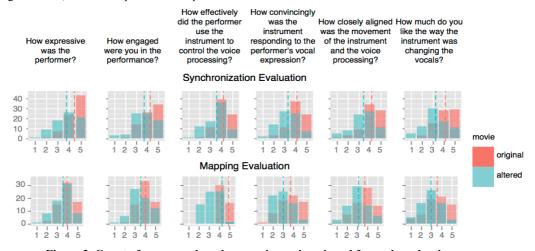

Figure 3. Count of responses based on movie version viewed for each evaluation.

"forceful" way. For example, one participant said: "It is not intuitive when a delay with a particular length lingers on; where the motion already changed to something else." Most of the participants noted the difference between gesture-vocal synchronization and desynchronization, as one participant described: "There is a natural relationship between the spinning motion and tension. Some effects work better than others."

One phenomenon is that, on nine occasions, participants gave a reason to pick an altered version as a preferred performance. While we cannot know for sure to what extent the processing influenced selecting a favorite out of the six performances, each mentioned how expressive the performer was. We might imagine that, in these instances, the performer's expressiveness outweighed the effects of our desynchronization. When participants preferred a performance that they heard in the synchronized condition, they tended to mention the instrument; otherwise, audiences attributed the expressiveness solely to the performer.

One song seemed to stand out as participants' favorite, especially in the synchronized version. The vocalist went into particularly high ranges and sang through several different octaves, and participants stated such things as "the different octaves presented interesting ways for the instrument to manipulate sound," and "the way it elongated the impressive high notes was what I enjoyed the most." This again highlights the perhaps complex relationship between how the instrument can accentuate the expressiveness of the performance, provided the performance is already detected to be expressive.

6.2 Intuitive Mapping Evaluation

As in the synchronicity experiment, subjects in the intuitive mapping experiment more frequently preferred the original performances and found them more engaging. On the other hand, participants' responses regarding the level of expressiveness of the performance did not significantly differ (though the result was trending) between the original and "nonintuitive" inverted versions. Participants further commented on expression in the free response, stating that the instrument added "something" to the performance; but it "just amplified what was already there." Two participants stated that the expressiveness of the performer remained independent. One said: "I think the instrument did not change the performer's ability to be expressive inherently, but did change the overall expressiveness of the music itself."

Conversely, some participants clearly noticed the relationship between voice processing and gesture intensity. For example, one participant said: "I liked when the rate of rotation of the instrument matched the speed (how short the chopped up audio bits were) of the effect. The relationship between the gesture and the effect seemed natural when intensity in the vocal performance aligned with intensity in the effects processing." In response to an inverted mapping video, one participant said: "The singer's long notes also sound longer in contrast to the quick moving murmuring effects." This further reflects how our original mapping between spinning gestures and vocal processing intensity is more intuitive.

6.3 Cross-Study Discussion

Notably, the perceived expression of the performer highly significantly differed between processing versions in Experiment 1, though it did not reach significance between mapping versions in Experiment 2. Since our question about the performance expressiveness asked the subject to only focus on the performance itself, which was in fact completely identical across all experimental conditions, participants not differing in response makes some sense in Experiment 2. However, from these results, we can only conclude that lack of synchronicity seemed to have a more marked effect on perceived performer expression than when the mapping was inverted, though further tests where all three versions are shown to participants can help disentangle this effect.

Across experiments, participants suggested that the instrument did not need to be used all the way through the performance. For example, one participant said: "I would consider not using the instrument the ENTIRE time you sing, maybe just do it at certain parts, I feel that will also add to dramatic effect." When not "overused", a participant felt "it was expanding the space and brought depth to her [the performer's] voice," and "it augmented the expression of the performance, providing additional dimensions of expression through the use of an instrument that looks analog and traditional, but that results in a very modern sound."

6.4 Experimental Methods Evaluation

Some improvements could be made to this experimental design and execution. First, bias may have been introduced in two manners, which may have influenced absolute ratings to the Likert-scale questions. For one, the performer in the videos was the person who conducted the experiment, which might have lead to an overall bias of the ratings. Bias may have also been introduced in the question phrasing – less bias may have been introduced were the questions phrased in a more balanced manner (i.e. "How convincingly *or not*…" rather than "How convincingly…").

However, since the purpose of the experiment was not to study absolute ratings, but rather the difference between one version and another, the potential for inflation does not seem wholly problematic. Furthermore, even if participants did detect that the manipulation was that some of the movies were synchronized with the gestures, while others were not; it was unlikely that they guessed it every time, and additionally, they did not explicitly know which version the study's authors were expecting to be preferred. Thus, the fact that the original version was more frequently preferred should not be invalidated by any contamination of the data or subjects attempting to "guess" the "preferred" answer. Bias introduced by response choices may also have been somewhat suppressed due to our "construct-specific" response choices (rather than agree/disagree response choices), which are often endorsed as a best practice in the questionnaire design literature [24].

Second, in the mapping experiment, we subjectively chose the single 'non-intuitive' mapping for comparison. Our results showed that the original mapping was more frequently preferred. However, if we had chosen a different mapping (from among infinitely many possible mappings) to compare, the results might be different. While this is a valid concern, testing more than one "non-intuitive" mapping was out of the scope of this study.

Finally, it is possible that some portion of the effects reported here may be due to differences in the audio alone, rather than audiovisual effects (e.g. one performance was more engaging than another simply because of how it *sounded*). Future studies using this or similar designs can include appropriate questions (i.e. those not related to movement) in evaluating audience response to the audio alone, and report how the inclusion of the video relates to any effects latent in audio differences.

7. CONCLUSION

We evaluated audience perception of the design of the TSPW. We selectively evaluated the TSPW's mapping from sensed horizontal spinning gesture (particularly the time-varying rotational speed) to vocal processing parameters. We designed two experiments to examine this mapping strategy's effectiveness in communicating musical expression from the performer to the audience, and to investigate our hypotheses about gesture-vocal processing relationships: synchronicity mapping and intuitive mapping. Our proposed evaluation methodology eliminated possible confounds associated with comparing different performances, and it had clear evaluation goals, stakeholders, criteria, methodology, and duration. We compared alternate gesture mappings for processing voice by processing sound from videoed performances in three different ways. Our analysis shows participants preferred our original

mapping (unaltered performance) across many metrics and found it to be more engaging. This work represents a first empirical evaluation of the TSPW and offers a framework for audience evaluation of electroacoustic vocal performance in the context of gesture-controlled vocal processing DMI.

8. FUTURE WORK

Future work could include user-centered design and usability studies for both expert vocalists and non-singers. For example, studying the perspectives of multiple vocalists asked to perform Western music pieces with the TSPW, as well as Tibetan monks or chanting practitioners whose feedback could shed light on the cultural context and help understand the possibilities that the TSPW could bring to both music performance and cultural exchange.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] J. Barbosa, F. Calegario, V. Teichrieb, and G. Ramalho. Considering Audience's View Towards an Evaluation Methodology for Digital Musical Instruments. *In Proceedings of the 2012 conference on New interfaces for musical expression*, 2012.

[2] J. Barbosa, J. Malloch, M. M. Wanderley, and S. Huot. What does "evaluation" mean for the NIME community? *In Proceedings of the 2015 conference on New Interfaces for Musical Expression,* pages 156-161, 2015.

[3] K.-E. Behne, & C. Wöllner. Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, 15(3), 324–342, 2011.

[4] G. Beller. The synekine project. *In Proceedings of the 2014 International Workshop on Movement and Computing*, ACM.

[5] C. Cadoz. Instrumental gesture and musical composition. *In Proceedings of International Computer Music Conference*, pages1-12, 1988.

[6] C. Darwin, P. Ekman, and P. Prodger. The expression of the emotions in man and animals. *Oxford University Press*, 1998.

[7] J. Davidson. What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, 6:279–301, 1994.

[8] M. Broughton, & C. Stevens. Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, 37(2), pages 137–153, 2009.

[9] C. Dobrian and D. Koppelman. The "E" in NIME: musical expression with new computer interfaces. *In Proceedings of the 2006 conference on New interfaces for musical expression*. IRCAM, Centre Pompidou, pages 277–282, 2006.

[10] Z. Eitan and R. Granot. How music moves: Musical Parameters and Listeners' Images of motion. *Music Perception: An Interdisciplinary Journal*, 23(3): 221–248, 2006.

[11] M. Schutz, & S. Lipscomb. Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36.6, pages 888–89, 2007.

[12] R. I. Godoy and M. Leman. Musical gestures: Sound, movement, and meaning. Routledge, 2009.

[13] D. Hewitt. Emic-compositional experiments and real-time mapping issues in performance. (2003): 96-104

[14] A. Johnston. Beyond evaluation: Linking practice and theory in new musical interface design. *In Proceedings of the 2011 Conference on New Interfaces for Musical Expression*, pages 280–283, Norway, 2011.

[15] S. Jorda and S. Mealla. A methodological framework for teaching, evaluating and informing nime design with a focus on expressiveness and mapping. *In Proceedings of the 2014 Conference on New Interfaces for Musical Expression*, pages 233-238, 2014

[16] P. N. Juslin. Music and Emotion, chapter Communicating Emotion in Music Performance. *Oxford University Press*, 2001.

[17] C. Kiefer, N. Collins, and G. Fitzpatrick. HCI methodology for evaluating musical controllers: A case study. *In Proceedings of the 2008 Conference on New Interfaces for Musical Expression*, pages 87-90, 2008.

[18] Küssner, B. Mats, D. Tidhar, H. M. Prior, and D. Leech-Wilkinson. Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in psychology* 5. 2014.

[19] T. J. Mitchell, S. Madgwick, and I. Heap. Musical interaction with hand posture and orientation: A toolbox of gestural control mechanisms. *In Proceedings of the 2012 Conference on New Interfaces for Musical Expression*, 2012. Ann Arbor, MI, USA, 21–23 May 2012.

[20] S. O'Modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):29–42, 2011.

[21] C. Pratt. The spatial character of high and low tones. *Journal of Experimental Psychology*, 13(3), 1930.

[22] S. Roffler and R. Butler. Localization of tonal stimuli in the vertical plane. *The Journal of the Acoustical Society of America*, 32(6):1260–6, 1968. un 1;43(6):1260-6.

[23] B. Rohrmann. Verbal qualifiers for rating scales: sociolinguistic considerations and psychometric data. (Unpublished manuscript) University of Melbourne, Australia, 2003.

[24] W. E. Saris, M. Revilla, J.A. Krosnick, & E. M. Shaeffer. Comparing Questions with Agree/Disagree Response Options to Questions with Construct-specific Response Options. *Survey Research Methods*, 4:61-79, 2010.

[25] W. A. Schloss. Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research*, 32(3):239–242, 2003.

[26] G.-M. Schmid. Measuring musician's playing experience: Development of a questionnaire for the evaluation of musical interaction. *In Practice-based Research Workshop at the 2014 Conference on New Interfaces for Musical Expression*, London, UK, 2014.

[27] B. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin. Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113, 2006.

[28] J. K. Vuoskoski, M. R.Thompson, E. Clarke, & C. Spence. Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, 76(2), 591-604, 2014.

[29] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from hci. Computer Music Journal, 26(3):62–76, 2002.

[30] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. Computer Music Journal, 26(3):11–22, 2006.

[31] J. C. Wu, N. Weitzner, Y. Yeh, J. Abel, R. Michon, and M. Wright. Tibetan singing prayer wheel: A hybrid musical-spiritual instrument using gestural control. In Proceedings of the International Conference on New Interfaces for Musical Expression, 2015.