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

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

ABSTRACT

This paper presents the first empirical evaluation of a digital music instrument for electroacoustic vocal performance. We study audience preference for the Tibetan Singing Prayer Wheel (TSPW); specifically, the way it maps horizontal spinning gestures to vocal processing parameters. We hypothesize that the levels of perceived expression and audience engagement increase when the mapping is (1) synchronized (such that the sensed gestures in fact control the processing in real time) and (2) intuitive. We filmed six songs with the singer simultaneously using the TSPW. In two experiments, two alternative soundtracks were made for each song. Experiment 1 compared the original mapping against a desynchronized alternative, Experiment 2 compared the original mapping (faster rotation causing a progressively more intense granular stuttering effect on the voice) to its inverse. All six songs were presented to two groups of participants, randomly choosing between alternate soundtracks for each song. Responses were evaluated via questionnaire. Viewers reported higher engagement and preference for the original versions, though level of perceived expression only significantly differed in Experiment 1. Through evaluating these control and mapping mechanisms, we aim to contribute to developing gesture theory in electroacoustic vocal performance, and designing DMI evaluation methodologies from different stakeholders' perspectives.

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 with a human computer interface that simultaneously processes vocals and synthesizes sound based on the performer's hand gestures with a one-to-many mapping strategy [29], introduced in NIME 2015.

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

The motivation behind the TSPW is to let the electroacoustic vocalist augment her/his 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 examined the input strategy of TSPW by measuring the system's latency, reliability, and reproducibility, as well as the 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.

2. EVALUATION BACKGROUND

A recent survey of evaluation within the NIME community [2] suggests that authors be clearer about evaluation goals, stakeholders, criteria, methodology, and duration.

2.1 Audience Perspective

Typical DMI evaluation case studies [13][16][23][30] 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 [26], 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 [18], enumerating several research studies that focused on a perceivable causal link between the gestural input and control mechanism and its produced sound [6][22]. 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 degree 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 and listener's perspective [14].

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" [18].

2.2 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 [3][12][17]. 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 [4][8][28][11], 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 [5]. Similar to daily speech, a vocalist's body movement and gestures are given more complex perceived cognitive meanings in terms of communicating emotions and expressing musicality to the audience as compared to other instrument players' [15].

The TSPW affords two main gestures: vertical (raising and lowering the device) and horizontal (speed of the wheel's spinning motion). The vertical parameter maps directly to the pitch of the synthesized sound by triggering various pitches and harmonic partials in the physically modeled singing bowls. This pitch-height relationship, in fact even the use of language that "low" means both "fewer cycles per second" and "closer to the ground," is supported by previous empirical work [19][21][24] and seen across a broad population, "applying to both musicians and nonmusicians' conscious and unconscious cognitive processes" [9]. Eitan and Granot further report that the listeners' perceived pitch contour is related to verticality.

Feeling on such solid conceptual ground with the vertical mapping, we focused our evaluation on the TSPW's horizontal spinning motion (i.e., the rotation of the "wheel"), which also controls aspects of the vocal processing. This gesture is not commonly found in DMIs and as a result, the evaluation of the audience perspective of this gesture and the kinds of processing that might be associated with it have not yet been studied. Thus, one of our primary evaluation goals was to validate the TSPW's designed horizontal spinning gesture control and mapping mechanisms, and their relationships with real-time vocal processing.

2.3 Audience Perspective Evaluation Methodology

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 [10][25]. Some of these studies' experimental design involves showing multiple versions of performances to audience members, who then provide judgments about the performances [7][10][31]. In some cases, videos of performers in which the same video of the performer is shown with audio manipulated, or the same audio is applied to multiple videos, to see the effect of gesture on auditory perception. For instance, Vuoskoski et al. [26] 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. Since the pianists were asked to perform three different versions, and the audio from one performance needed to be synced to the others, time-warping algorithms were used on the video to convincingly match the synchronization of the audio and the video.

Our study takes a similar approach of using a consistent stimulus across conditions. Whereas in the case of Vuoskoski et al., the differing source material was the multiple expressive conditions of the pianists' performances, in our study, different versions of processing of the same raw vocals are created, resulting in a differing processing, but yet still the fundamental timing of the performer's motions (her singing resulting in time-synchronized sound) line up with the video.

3. RESEARCH QUESTIONS

3.1 Synchronicity of Processing

Within the NIME field, where the researchers build instruments, it is 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 Defining "Intuitive Gestural Mapping"

Generally in instrumental performance contexts, faster and/or larger motions produce faster and/or more energetic sound. For example, the relationships between music intensity and the performer's motion are studied from the listener's' perspective by Repp [20]. Repp relates this finding to "kinematic implications of musical structure, which induce perceptual biases in listeners". Repp argued that the perceptual sound-motion relationship may be understood as perceptual-motor interaction.

Therefore we define "intuitive gestural mapping" as the original TSPW's mapping strategy, in which faster horizontal spinning hand motions produce faster and more intensely granulated vocal processing effects, and vice versa.

3.3 Hypotheses

Between two experiments: the "synchronization evaluation" experiment (Experiment 1) and "intuitive mapping evaluation" experiment (Experiment 2), we have 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.

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

4. METHDOLOGY

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. The duration of the entire evaluation design and implementation required half a year, including experimental design, human subjects recruitment, pilot studies, method adjustments, formal experiments, data collection and analysis, etc.

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" [31]). The intuitiveness-evaluation instantiation 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., vocalist, piano, guitar, flute, drum set, cello, trumpet, oud, harp, gamelan, and drums). All participants were entered into a lottery for one of five $20 gift cards.

4.2 Musical Stimuli

Six vocal pieces of music (1-2 minutes each M = 1.34 min, SD = 30 sec) were composed for the prayer wheel. The vocal pieces represented a variety of styles and affects (rhythmic, melodic; four of which were in a Tibetan/Nepali style of song, one was a Chinese folk song, and one was a Chinese spoken word piece), and were videotaped and audio recorded by a vocal performer fluent with the instrument (the first author) to eliminate potential variability in perceived expressiveness of different performers. In each one of the experiments, two versions of the processing were made: the original vocal processing [29], and a separate processing designed to test the study's hypotheses, described in session 2.3.

The dry voice and the changing values of the two parameters controlled by the horizontal spinning gesture, namely pulse width and pulse interval, were recorded as time-synchronized wav files. These were input to a modified PureData patch to reprocess the voice in desynchronized and reversed parameter versions.

4.2.1 *Synchronization Evaluation Stimuli*

For the *"desynchronized"* version, the gestures and the vocal processing are not tightly aligned with each other. Each of the six performances produced a unique pattern of TSPW rotational speed over time, mapped to control the pulse width and interval parameters in a delayed and windowed voice processing algorithm [29]. The pulse width was inversely related to speed, and the pulse interval was directly proportional. To create a desynchronized TSPW sound, we reprocessed each dry voice recording using the recorded pulse width and interval of a different performance. For example, the dry voice of composition 1 was reprocessed with parameter envelopes taken from the performance of composition 2, the dry voice of composition 2 was reprocessed with parameter envelopes from composition 3, etc. The effect is to apply a final sequence that is uncorrelated with the original sequence, but with the applied time-varying control functions qualitatively similar to the "correct" control functions from the original performance (being generated with the same instrument and by the same performer in similar performance contexts). For example, Figure 1 shows the pulse width control parameter as a function of time for compositions 2 and 3.

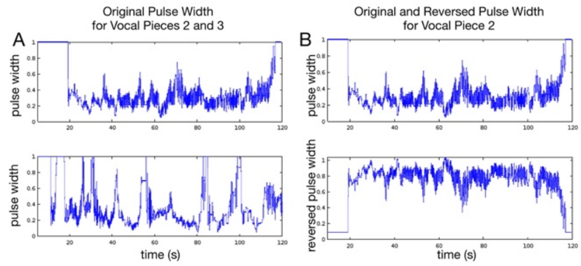

Figure 1: A) Prayer Wheel-controlled pulse-width values used to process dry voice in vocal piece #2 (top) and #3 (bottom). B) Original prayer wheel-controlled pulse-width values (top) and the corresponding Reversed pulse-width values (bottom) for vocal

4.2.2 *Mapping Evaluation Stimuli*

For the "reversed" version, a flipped version of the audio processing mapping is applied to the dry voice. The pulse width and interval parameters first were normalized, then subtracted from 1, effectively making previously small values large, and vice versa. In this reversed version, faster spinning hand gestures map to a longer pulse and thus a slower and longer reverb.

Creating the reverse 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, the performer kept the TSPW in motion throughout all 6 pieces except for only a few complete stops. The normal TSPW usage pattern skews towards higher spinning speeds, with limited stopped or zero values, while the reversed TSPW parameters rarely hit the capped speed, and contain longer periods at zero value.

Reversing the parameters thus created a highly reverberant processed sound that could not be produced with a real TSPW. To keep the reversed version from sounding like a qualitatively different instrument, we experimented with several offsets before selecting a 0.09 offset added to the reversed parameters. Figure 1B compares the original and reversed 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 video was the desynchronized version, and for experiment 2, the altered video was the 'inverted' version. Each participant then randomly watched either the original or altered for each of the performances, so he or she never viewed both versions of a piece. This design was chosen so that, in the aggregate, there was an approximately equal number of viewings of each version of each performance, and allowed for 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 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 to perform, participants were instructed to rate the performances and to pick their favorite ones based on the vocal processing method and how the instrument was used during the performances. 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 six 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?", "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." An entire experimental session took 20-30 minutes.

5. RESULTS

5.1 Likert-Scale Questions

Since the Likert-scale questions resulted in ordinal 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 ($p < .001$ always). 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 ($p < .05$ always), while the response to Q1 was nearly so ($p = .07$). Figure 2 shows the count of responses based on movie version for each condition of the two evaluations. Group means are presented as dashed vertical lines. The red band above each "5" indicates that subjects viewing the original version of the mapping were always more likely to answer "extremely."

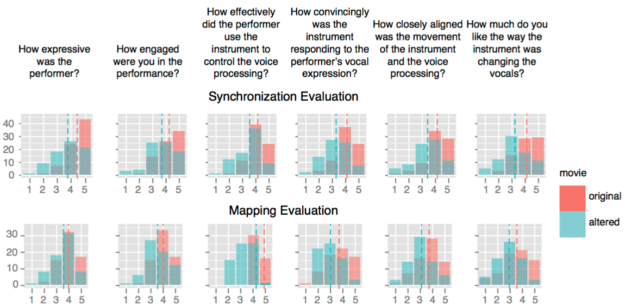

Figure 2. Count of responses based on movie version viewed for each evaluation.

5.2 Free-response Question

The final question of the study 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 one video - we have included each reason discussed for each video mentioned in the table.

Table 1. Counts per movie type of various categories of free-response answers to why a particular performance was preferred by the participant.

	Synchronization		Mapping	
Response type	Original	Altered	Original	Altered
Instrument added to performance, instead of distracted 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 match 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

Notably, in the synchronization evaluation, the intuitive matching of the gestures, and the vocals present in the original version was a frequent reason given for why a video was preferred, and in no instance did a person refer to the altered videos as being more intuitive. Interestingly, there was an equal number of 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 their favorite.

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.

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 [27]. Our straw man non-synchronous mapping, a situation where the movement does not affect the sound processing, was (not surprisingly) tended to be less favored by audience members.

In order to further examine this assumption, we asked six questions of our participants relating to perception of performer expression, participant engagement, and perception of the relationship between movement and processing. 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 "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 expressiveness of the performer was more important than 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

In the synchronicity experiment, people more frequently preferred the original performances and found them more engaging and expressive than the altered ones. However, in the intuitive mapping experiment, while people still tended to prefer the original and find it more engaging, people generally didn't find the performance to be more expressive in the original than in the altered versions ($p = .07$). For example, several participants stated 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." When using the alternative mapping where the gestures and vocal processing intensity are in the reverse relationship, one participant said: "The singer's long notes also sound longer in contrast to the quick moving murmuring effects." This reflects the data showing that a more intuitive mapping between spinning gestures and vocal processing intensity results in a more highly engaged audience and a performance perceived to be more expressive.

6.3 General Discussion

Some improvements could be made to this experimental design. First, the performer in the videos was the person who conducted the experiment. This might have lead to an overall bias of the ratings. However, since the purpose of the experiment was not to study absolute ratings to the questionnaire, but rather the difference between one version and another, the potential for inflation does not seem wholly problematic. Further, even if participants did detect that the manipulation was that some of the movies was 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 the one that was more frequently preferred should not be invalidated by any contamination of the data or subjects attempting to "guess" the "preferred" answer.

Second, in the mapping experiment, the single 'non-intuitive' mapping for comparison was subjectively chosen by us. Our results showed that this was more frequently not preferred to the original. However, if we had chosen a different mapping (from among infinitely many possible mappings) to compare, the results might be different. While a valid concern, testing more than one "non-intuitive" mapping was out of the scope of this study.

Third, 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, the participants are "correct" not to perceive a difference. However, there is still a perceived difference of the performance's expressiveness across the synchronized/unsynchronized conditions. It appears that some participants can not clearly comprehend our questions in the way that we expect them to. Therefore, how to precisely articulate and phrase our questions is crucial if we expect the audience to give us the most useful and accurate feedback.

Last but not least, regarding the method of how to use this instrument in general, audiences suggest that the instrument does 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."

7. CONCLUSION

We evaluated audience's perception of the design of the TSPW. We selectively evaluated the TSPW's mapping between vocal processing parameters and sensed horizontal spinning gesture, in particular the time-varying rotational speed. We designed two experiments to examine this mapping strategy's effectiveness in communicating musical expression from the performer to the audience, and to prove 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 subjects preferred our original mapping (unaltered performance) across many metrics and found it to be more engaging. This is the first empirical evaluation of a DMI for augmenting electroacoustic vocal performance from the audience's perspective. It offers a framework for audience evaluation of electroacoustic vocal performance in the context of gesture-controlled vocal processing DMI.

8. FUTURE WORK

The future work could be user centered design and usability studies for both expert vocalist and non-singers. For example, asking multiple vocalists to perform western-music pieces with it, and study their perspectives. Alternatively, providing the TSPW to Tibetan monks or chanting practitioners and letting them use the TSPW to practice their traditional chanting, and study their feedback could also be interesting in the cultural context, and to better understand the possibilities that the TSPW could bring to both music performance and cultural exchange.

9. ACKNOWLEDGMENTS

Our thanks to the original development team of the TSPW instrument. Thanks to UCSB/CREATE and Curtis Roads for the experimental facilities and funding support. Thanks to the UCSB MAT writing club. Thanks to Stanford/CCRMA and Chris Chafe for the experimental facilities support.

10. REFERENCES

[1] J. Barbosa, F. Calegario, V. Teichrieb, and G. R. P. McGlynn. Considering audience's view towards an evaluation methodology for digital musical instruments. In Proceedings of the International 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 New Interfaces for Musical Expression, pages 156-161.

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

[4] C. Cadoz. Instrumental gesture and musical composition. In Proceedings of International Computer Music Conference, 1988.

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

[6] 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.

[7] Broughton, M., & Stevens, C. (2009). 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), 137–153.

[8] C. Dobrian and D. Koppelman. The "E" in NIME: musical expression with new computer interfaces. In Proceedings of New Interfaces for Musical Expression, pages 277–282.

[9] Z. Eitan and R. Granot. How music moves. Music Perception: An Interdisciplinary Journal,, 23(3):221–248, 2006.

[10] Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences per-ceived tone duration. Perception, 36, 888–89

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

[12] D. Hewitt. Emic-compositional experiments and real-time mapping issues in performance.

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

[14] 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 International Conference on New Interfaces for Musical Expression, 2014.

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

[16] C. Kiefer, N. Collins, and G. Fitzpatrick. Hci methodology for evaluating musical controllers: A case study. In Proceedings of the International Conference on New Interfaces for Musical Expression, 2008.

[17] 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 International Conference on New Interfaces for Musical Expression, 2012.

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

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

[20] B. Repp. Musical motion in perception and performance. Timing of behavior: Neural, psychological, and computational perspectives, pages 125–44, 1998.

[21] 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.

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

[23] G.-M. Schmid. Measuring musician's playing experience: Development of a questionnaire for the evaluation of musical interaction. In Proceedings of the International Conference on New Interfaces for Musical Expression, London, UK, 2014.

[24] O. Trimble. Localization of sound in the anterior-posterior and vertical dimensions of auditory space. British Journal of Psychology, 24:320–334, 1934.

[25] 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.

[26] 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.

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

[28] T. Winkler. Making motion musical: Gesture mapping strategies for interactive computer music. In Proceedings of International Computer Music Conference, 1995.

[29] J. 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.

[30] Q. Yang and G. Essl. Evaluating gesture-augmented keyboard performance. Computer Music Journal, 38(4):68–79, 2014.

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