2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) | 979-8-3503-2745-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACIIW59127.2023.10388113

Vocalization for Emotional Communication in Crossmodal Affective Display

1st Pranavi Jalapati
Department of Computer Science
University of Southern California
Los Angeles, United States of America
jalapati@usc.edu

2nd Selwa Sweidan
Department of Cinematic Media Arts and Practice
University of Southern California
Los Angeles, United States of America
sweidan@usc.edu

3rd Xin Zhu
Department of Computer Science
University of Southern California
Los Angeles, United States of America
xinzhu@usc.edu

4th Heather Culbertson
Department of Computer Science
University of Southern California
Los Angeles, United States of America
hculbert@usc.edu

***Abstract*—This paper presents our design of a crossmodal vocalization-haptic system to allow users to communicate emotions to a partner or pair of users. We explore affective context as a combination of user relationships (specifically the closeness between pairs of users) and user culture. We share the design and implementation of the crossmodal system that takes up to ten seconds of vocal expression (including humming or singing) from one user and transposes it into haptic signals to be displayed to twelve vibration actuators worn on the forearm of the second user. Our method of transposing musical vocal inputs captures the key signal features of rhythm, amplitude, time, and frequency. We present the results from a human subject study (N=20) involving 10 pairs of users with varying levels of closeness (ranging from siblings, friends, and strangers) to understand how our system supports affective communication. Our results show that low-level and rhythm audio parameters most strongly influence affective responses in our users. Additionally, the low-level vocal features are influenced by user demographics and the closeness between the pairs of users. The results suggest the impact of user closeness on affective communication and provide insights into the optimal music transposition methods best suited for affective communication.**

***Index Terms*—musical haptics, crossmodal, affective communication, user personalization**

I. Introduction

This paper contributes to understanding the ways in which vocal and haptic signals can be used to communicate affective information. We offer an approach to building and testing affective user-to-user systems that allows for subjective, and personalizable forms of affective communication. Our challenge has been to create a system that can transpose affective communication from sound to touch, while also supporting user-to-user expression. Working with sound (as the mode of expression from the user sending the signal) and haptic vibration (felt by the recipient user receiving the transposed signal), underscores our interest in the affective potentials for crossmodal communication. Through our design and user study, we explore whether emotionally close user pairs (family, friends, and romantic partners) are able to share communication that is affectively stimulating. Our research explores what it means to transpose socially affective and personalizable communication between users. We explore affective context as a combination of user relationships (i.e., meaning the closeness between user pairs) and user culture.

A key research question focuses on whether affective aspects of vocal expression as an input, can be transposed and mapped to haptic output between pairs of users. This builds on research that shows that haptic cues, or touch sensations, can play a role in communicating emotions. Haptic cues have been shown to convey emotional information by modulating the intensity, duration, and location of touch. For example, gentle touch on the arm increased the perception of trust and decreased the perception of anger in an interaction between strangers [1]. Additional research shows that touch can be a powerful emotional communicator, particularly in close relationships, where it can convey affection, comfort, and support [2]. Research has also shown that haptic cues can be used to enhance emotional expressions in virtual environments. For example, incorporating haptic feedback into virtual hugs significantly increased the experience of social interaction and emotional involvement in users [3].

It is important to note that the influence of context on emotion perception is significant. Discrete emotional meanings are contingent upon the context in which they are experienced [4]. Additionally, the effectiveness of haptic cues in communicating emotions can also be influenced by individual differences, such as culture, gender, and personal experiences with touch [5]. All of these point to how haptic cues or touch sensations can effectively communicate emotions, but also show that this effectiveness is influenced by aspects such as the specific touch, the context of the touch, and individual differences [6].

Vocal expression through humming or singing is a form of affective communication that has been studied in the fields

979-8-3503-2745-8/23/$31.00 ©2023 IEEE

of communication theory and affect theory. Within the field of communication theory, humming and singing have been considered forms of ”paralanguage”, or nonverbal vocal cues that accompany speech and carry meaning beyond the verbal content of words. Paralinguistic cues, such as humming or singing, have been found to convey affective meaning and contribute to the overall communicative intent of an interaction [7]. Within affect theory, humming and singing can be used to communicate emotions in a manner that is abstracted from the verbal content of words, providing a unique and powerful form of emotional expression [8]. This means that our system is designed for vocal inputs (humming and singing) which are already abstracted, or far from having a fixed functional meaning or interpretation, and are highly subjective in interpretation and affect.

Another factor for exploring user-to-user affective communication is related to ethical concerns more broadly in the field of affective computing. Traditional approaches to affective computing have been criticized for recognizing affective states and exploiting such recognition to make inferences and shape human decisions [9]. Affective computing scholars argue that even when affective systems seem harmless, the use of emotional manipulation may be problematic [9]. In fact, researchers have gone as far as sharing the recommendation to “ban the use of affect recognition in important decisions that impact people’s lives and access to opportunities” [10]. While our systems focus on the potential of affective communication, we do not exploit the recognition of affective states. Instead, our design offers pairs of users the agency to invent and share idiosyncratic communication without a system making inferences and offering responses to such exchanged data. While much of our testing focuses on gaining insights about which audio parameters most strongly influence affective responses in our users, our audio analysis supports the decoupling of matching or limiting user potential through semantic matching of emotion and valence, to instead focus on the musical potentials of affective communication and focus on transposing these to effective stimulus for affective communication.

Affective haptics builds off of decades of research and development of affective computing systems that can recognize, process, and represent human emotions [11]. Our research focuses on designing for haptic personalization and user-to-user communication. In designing for user-to-user communication, the user is given the agency to idiosyncratically generate affective communication, without a system stabilizing or assuming “correct” modes of affective communication. While our human subject study attempts to understand whether emotion and valence are communicable in our cross-modal system, our system does not assume any matching or predefinition of emotion, valence, or any aspect of affective communication. Our research contributes to affective haptics research and crossmodal musical haptics (from sound to touch), which allows for idiosyncratic and highly personalizable transmissions between users.

The contributions of this work include:

- A novel end-to-end architecture to capture raw audio input, convert it to a haptic signal, and transpose it to a physical medium.
- Enable affective communication using haptics.
- Haptic algorithm for transposing sound signals to haptic signals to better capture affect.
- Insights into how context and emotional closeness influence affective communication strategies.
- Results suggest the relationship between context between participants, vocal communication, and affective communication.

II. Background

Our design contributes to the emerging interdisciplinary and intersecting fields of affective haptics and musical haptics. Musical haptics investigates touch and physical feedback in musical experiences. It is often used to create new musical instruments and devices with haptic feedback. The field explores a range of applications including musical instruments, which provide additional sensory feedback to the performer, or enhance the overall musical experience [12].Applications also include music therapy, to help patients with various physical and emotional conditions, such as Parkinson’s disease or anxiety [13], and also music education [14]. Musical haptics has also been applied to virtual and augmented reality, to create immersive musical experiences and interact in a more tangible way [15].

Affective touch as a form of communication, can modulate emotions [16], [17], and “increase trust and affection” [18]. Further, social touch “intensifies the perceived social presence of remote communication partners” [18]. Indeed, storytelling is enhanced when accompanied by touch [19], which is an important consideration for affective haptics which focuses on designing for affective and interpretive communication, rather than informational accuracy. Further, a “large part of interpersonal communication is emotional rather than factual” [20], which supports our approach in seeking to communicate affective information between users.

Affective haptics builds off of decades of research and development of affective computing systems that can recognize, process, and represent human emotions [11]. Our design focuses on user-to-user communication which allows us to consider how to design for user agency. In designing for user-to-user communication, the user is given the agency to idiosyncratically generate affective communication, without a system stabilizing or assuming “correct” modes of affective communication. In other words, our system for displaying affective haptics from vocal expression does not assume a stable haptic lexicon. To achieve this, it is important to consider to what degree our approach for creating haptic sequences (as transpositions of vocal inputs) can offer discernible and differentiated [21] affective impressions between users.

Since we are exploring the context, it is important to consider how culture is a factor in understanding emotions, and how culture plays an essential role in fundamental processes such as labeling and defining our emotional expe-

riences [22]. Specifically, non-linguistic vocal expression is influenced by cultural factors, including cultural variations in vocal expressions, cultural display rules, perception and interpretation of vocal expressions, cross-cultural communication challenges, and cultural adaptation. Research has demonstrated that non-linguistic vocal expressions differ across cultures. For example, studies have highlighted variations in laughter styles, crying patterns, and vocalizations associated with specific emotional states, reflecting cultural norms, values, and socialization processes [23]. Specifically, while the non-verbal vocalization of predominantly negative emotions exhibits cross-cultural recognition, in contrast, the expression of most positive emotions relies on culture-specific signals [23]. Understanding these influences is crucial for considering how we may design and study affective communication between users.

To design a haptic experience that satisfies rather than intrudes, one approach is to actuate vibrations by varying rhythm and frequency [21]. Haptic designers and engineers have correlated auditory and vibrotactile stimulation for music enhancement [24]. A recent study [24] exposed participants to melodious sounds through supra-aural headphones and vibrotactile stimuli through haptic actuators [24]. Specifically, actuators were triggered by mapping frequency, intensity, and timing with musical features like pitch, intensity, and rhythm [24]. Another haptic project designed for cochlear implant users takes audio signal processing and makes it more dynamic by customizing the intensity levels based on the positioning of the band and detection thresholds of the user [25]. This offers insight into ways that audio-to-vibration transposition can be mapped effectively and informs our choice to vary the signal intensity of vibration to correspond to the musical octave that the user inputs. Another important parameter when designing for vibrotactile communication is that users have five to ten seconds of tactile memory [21], which has informed our choice to limit user vocal input to a maximum of ten seconds.

Our design also builds on research for algorithmic detection of sound and emotion. This active area of research in the field of affective computing focuses on developing systems that can automatically recognize and categorize emotions expressed through sounds such as speech and music [26]. One approach for detecting emotions in speech is to analyze acoustic features such as pitch, loudness, and speaking rate [27]. However, emotion recognition in music is considered more challenging due to the complexity of musical signals and the subjectivity of emotional responses to music [8]. Researchers have explored the use of audio features such as rhythm, melody, and harmony to classify emotions in music [28].

III. Methods

A. System

Our device is designed for two users to share crossmodal affective communication, with one user sharing up to ten seconds of vocal expression (such as humming or singing) and a second paired user receiving the transposed signals as haptic

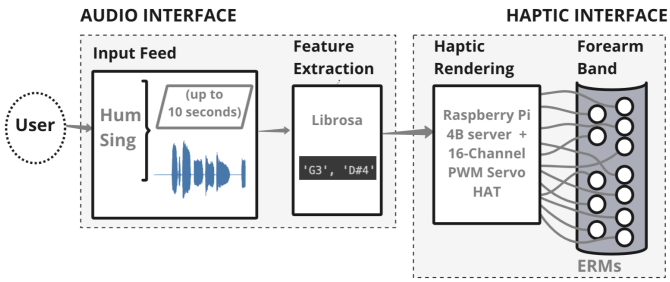

Fig. 1. Flow diagram showing data transmission from recorded audio to displayed vibration.

output. Features of the vocal expression (including rhythm, amplitude, time, and frequency) are transposed into haptic signals displayed across twelve vibration actuators worn on the forearm of the second user. Our cross-modal affective communication system, shown in Fig. 1, consists of two distinct subsystems: an audio interface and a haptic interface.

1) Audio Interface: The audio workflow starts with a microphone used to record up to 10 seconds of audio from one user singing or humming. This file is then processed using the `Librosa` Python library for music and audio analysis. We estimated the pitch of the recorded audio using a probabilistic Yin (pYin) model [29] with a window length of 1024 (which is ˜23 mS when sampled at 44100 Hz). We used the Viterbi algorithm to determine the transition from one note to another [30]. The algorithm takes in a transition matrix with the probability of going from one state to another (source to destination) and a transition matrix with the prior probability related to the source state at a given discrete time. Using a two-state acoustic model with the onset and sustain phases, the required parameters such as onset (start time), sustain (end time), note, and pitch are generated for each frame being analyzed. These values are recorded in a CSV file generated for each audio sequence; the file is saved to be used by the haptic interface. For acoustic verification, a midi file is generated for the transformed notes. The raw audio file, a pre-processed audio file, a midi file, and the CSV file for audio-to-note mappings are saved in the process.

2) Haptic Interface: The transposed audio signal is displayed as haptic feedback through a sleeve equipped with eccentric rotating mass (ERM) vibrotactile actuators. The design of the band is inspired by the heptatonic scale (conventionally called the Western scale) in C Major. The sleeve has 12 actuators in total, 7 for each music note (C, D, E, F, G, A, B) and 5 for the equivalent sharp notes. An alternate approach would have been to use 7 actuators each playing both the note and the sharp note Fig 2. But this would result in discontinuous perception and the sharp-note to note extrapolation would make it difficult to relate the haptic perception with the audio played. The haptic rendering algorithm runs on a Raspberry Pi 4B server and a 16-Channel PWM Servo HAT for Raspberry Pi is used to ensure adequate power supply to the actuators.

The underlying haptic rendering algorithm iterates through

Fig. 2. Image of the sleeve with 12 ERM actuators

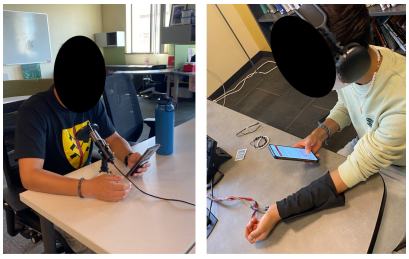

Fig. 3. Experimental setup showing (left) audio recording and (right) haptic display. Participants were given an affective prompt and provided survey responses on a smartphone.

the rows in the music sheet (generated by the audio interface) and activates the actuator that corresponds to the current note. For each row, if the onset time (start_time) of a note has occurred or passed, then the actuator for the corresponding note is activated for the given duration (end_time).

A pilot study conducted showed that the most common octaves for non-professional singers are 2, 3, and 4 (that is, notes ranging from C2 to B4). Given that most of our users are not professional singers, the recorded audio did not comply with the rules of music theory and the notes ranged all the way from the 2nd octave to the 4th octave in a single file. The key difference between the octaves is the frequency where each octave plays at a frequency that is a multiple of the fundamental frequency F0. The challenge with simulating different octaves by changing the frequency is there is no perceivable difference for different frequencies. Hence we created our system analogous to audio perception by adjusting the strength of vibration for the actuators. A lower octave (bass with lower pitch) sounds sharper while a higher octave (treble with higher pitch) sounds blunter. We vary the strength of the vibration to display pitch using pulse width modulation (PWM); a higher duty cycle is used for a higher octave, which increases both the strength and frequency of the vibration for higher octaves. The PWM duty cycle was set to 75% for the 4th octave, 50% for the 3rd octave, and 25% for the 2nd octave. The benefit of using multi-threading for running the actuators is that it ensures the continuity of signal and also helps in simulating the attack, decay, sustain, and release phases (ADSR model) [31] of the music envelope.

B. Experimental Methods

Our study explores affective communication through cross-modal signals (from sound input to haptic output) between pairs of users. Specifically, our study considers whether users are able to associate the rendered haptic signal as an emotional stimulus.

We recruited pairs of participants who self-identified as friends and romantic partners, and strangers, through open calls shared via email lists, departmental email announcements, and word-of-mouth invitations. This study was approved by the University of Southern California's Institutional Review Board (IRB) under Protocol UP-20-01131, and all participants gave informed consent. Ten pairs of users participated in the study for a total of 20 participants (9 female, 10 male, 1 non-binary; ages 18-44). Two participants self-disclosed that they were sisters, 12 participants were friends, and 6 participants were strangers. The total duration of the user study was one hour including a five-minute break. Of our 20 individuals comprising 10 user pairs, 4 pairs were from the same country of origin (and actually sang to each other in non-English language) and 6 pairs were from different countries of origin. One pair identified as both being from China and three pairs self-identified as being from the USA. The rest of the 6 pairs were not from the same country of origin.

Before starting the study, all participants were asked to complete a pre-survey with demographic questions, the Comfort with Interpersonal Touch (CIT) questionnaire [32], and the Inclusion of Self in Others scale [33].

The study was split into two phases and in both phases, one participant was asked to vocalize (sing or hum) for 10 seconds based on an affective prompt. This vocalization was recorded and processed using the steps described above. The second participant received the cross-modal stimulus including a haptic stimulus through the array of haptic actuators worn on their forearm and an audio stimulus by playing the recording from the first participant. The order of cross-modal stimuli was either haptic first (HF) where the haptic sequence was played first followed by the audio sequence, or audio first (AF) where the audio was played first followed by the haptic sequence, or simultaneously display (SD) where both are played in parallel. The choice of this ordering was randomly chosen for each clip played for the second participant.

After a short break, the participants switched roles in the second phase of the study where the vocalizer wore the sleeve and experienced the cross-modal stimulus. To ensure that the displayed sensation is not influenced by the noise leak from the vocalizer or other background noises, participants were physically separated from one another, and the participant receiving the haptic feedback wore noise-canceling headphones. The workflow is completely asynchronous, and both users were asked to fill out a survey after each sequence.

In Phase I, the first participant (vocalizer) was given five randomized emotional prompts ("excited", "happy", "neutral", "sad", and "angry") and was asked to sing or hum vocal sequences. These prompts were chosen in order to provide a set of five emotions with a wide range of both valence and arousal. The recordings were then processed by the audio interface and the musical notations are generated. The

TABLE I
SURVEY QUESTIONS FOR PARTICIPANTS PROVIDING AUDIO INPUT AND RECEIVING HAPTIC FEEDBACK

Question List	
Question 1	Did you feel any emotion? - Selected Choice
Question 2	Rate the strength of your emotion: - How strong is your emotion?
Question 3	What emotion do you think is being conveyed?
Question 4	Rank the emotional strength of the vocal expression that you just heard: - How strong is the emotion in the vocal expression?

notations were then uploaded to the haptic interface to generate the haptic mapping to be played on the sleeve worn by the second participant (sleeve wearer).

After a short break, the participants swapped roles for Phase II. The process adopted for Phase II is the same as Phase I with the exception of using three valence prompts (”positive”, ”negative”, and ”neutral”) instead of 5 emotional prompts. One thing to note is that in our study we gave descriptive prompts (ranging in valence and emotion) to our user testers who vocalized expressions. However, in a real-world scenario a user would not receive an external prompt, and instead express something at will, based on their own affective state. However, for our study, we gave descriptive prompts to compare results within pairs of users, and across all pair users. With this hypothesis that receiving an expressive prompt differs from what our systems' real-world context would allow for, we attempted to vary the prompts between each phase. In Phase I the prompts were more concrete (“excited”, “happy”, “neutral”, “sad”, and “angry”), relative to the prompts in Phase II, which were broader (”positive”, ”negative”, and ”neutral”). Table I shows the survey questions for participants receiving haptic and audio feedback.

IV. RESULTS

For this analysis, we focus on the ability of our system to convey valence and arousal rather than distinct emotions. Therefore, we first separate the emotional prompts of Phase I into discrete categories of valence and arousal. Emotions ”excited” and ”happy” are labeled as positive valence, and ”sad” and ”angry” are labeled as negative valence. Similarly, ”excited” and ”angry” are labeled as high arousal, and ”happy” and ”sad” are labeled as low arousal. ”Neutral” is labeled as neutral valence and neutral arousal.

A. Audio Analysis

The audio recordings from the vocalizer were analyzed to gain a deeper insight into the audio features that constitute the affective component of communication. Based on the findings in [34], low-level and rhythm are critical features required for determining the audio signal's arousal, and low-level and tonal features are critical for determining the signal's valence. Low-level features are statistical features extracted from audio. Rhythmic features describe the stylistic features of audio such as the pattern of beats. Tonal features refer to the melody and harmony of audio through features like pitch and chords. To gain further insights, the Essentia library[1] was used to extract the different low-level, rhythm, and tonal features of audio. We then conducted a series of one-way ANOVA tests below on each of these features to identify their significance in capturing the valence and arousal of the conveyed emotion.

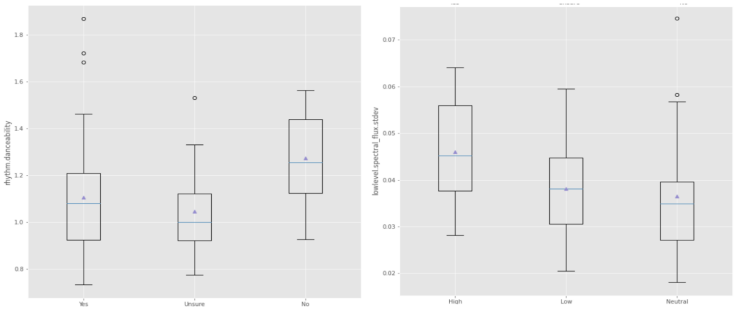

Fig. 4. Distribution of (left) danceability vs perceptability and (right) standard deviation of the spectral flux vs the vocalizer's emotion

1) Features for Arousal: To study the relationship between arousal and the audio parameters, the low-level and rhythm features were analyzed using a set of one-way ANOVAs with arousal as the factor and various audio parameters as the dependent variable. The results suggest that the signal's danceability feature [35] has a significant influence ($p = 0.046$) on the perceivability of the haptic sequence. Danceability is a mixed measure of tempo and beat strength and suggests the ease with which a person can dance to audio. Also, arousal of the emotional prompt given to the vocalizer has a significant influence on the standard deviation of the spectral energy flux ($p = 0.031$) feature of the audio sequence generated. Spectral flux is a measure of how the relative magnitudes of frequency are changing across audio frames. Fig. 4.

2) Features for Valence: To study the relationship between valence and the audio parameters, we conducted a set of one-way ANOVAs with valence as the factor and various low-level and tonal features [36] as the dependent variable. The results

[1]https://essentia.upf.edu/

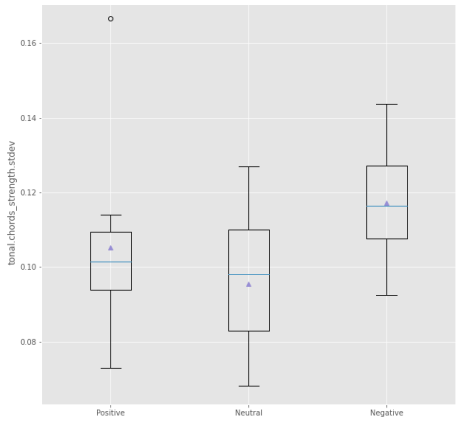

Fig. 5. Relationship between the vocalizer's emotion and the standard deviation of the chord strength.

suggest that the emotional prompt given to the vocalizer has a significant influence on the standard deviation of the strength of the tonal chords ($p = 0.023$) feature of the audio sequence generated as shown in Fig. 5. The strength of tonal chords is indicative of the likelihood of the tone's associated chord progression. There are no significant correlations with low-level features ($p > 0.05$).

B. Data Analysis

Our study examines the effect of cross-model affective communication by changing the stimulus sequence. For Phase I, we first analyze the ability of the second user to correctly understand the valence of the interaction. We do not focus on analyzing the accuracy of the conveyed emotion itself because Fisher's test on emotion between different orders of stimulation did not show any significant difference in participants' ability to receive emotions ($p > 0.2$). In this phase, the haptics first (HF) conditions were conducted 20 times and 16 of them correctly matched valence (80%), audio first (AF) was conducted 20 times and 8 of them matched (40%), and simultaneous display (SD) (haptics and sound are played at the same time) was conducted 10 times of which were 5 correct (50%). Given the smaller sample set, we ran a Fisher's exact test to determine if there was a significant association of affective communication accuracy between the order of provided feedback. This analysis showed that participants were significantly more accurate at determining the conveyed valence during the HF condition than the AF condition ($p = 0.022$). There was no difference between the accuracy of AF and SD ($p = 0.705$) or between HF and SD ($p = 0.115$).

We also evaluated the accuracy of recognizing the valence of the interaction in Phase II. In this phase, the HF condition was conducted 10 times of which 4 were correct, the AF was conducted 11 times of which all 11 were correct, and SD was conducted 9 times of which 3 were correct. We ran a Fisher's exact test to determine if there was a significant association between valence accuracy and the condition order in this phase. This analysis showed that participants were significantly more accurate at determining valence in the AF condition than both the HF ($p = 0.003$) and SD $p = 0.002$). There was no significant difference between HF and SD ($p = 1$).

We additionally studied the effect of crossmodal communication as impacted by demographics and closeness. In particular, we looked at vocal patterns between pairs of users from the same and different countries of origin and also the closeness of users and found a correlation between spectral flux (a low-level audio feature) and the users' continent of origin[2], which are shown in Fig. 6.

[2]We acknowledge the process of grouping countries and how politically charged such a designation can be.

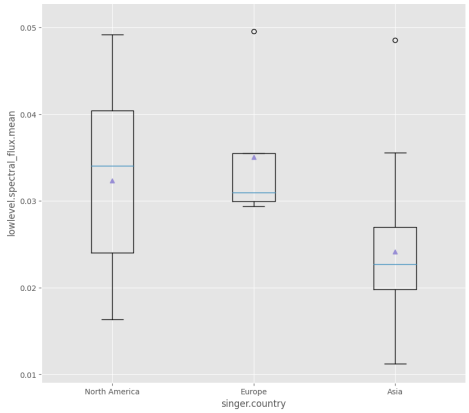

Fig. 6. Relationship between the mean of the spectral flux vs the users' continent of origin.

V. Discussion

A. Analysis

We examine the effectiveness of using a crossmodal method at communicating affective information with our system. Our results show that participants are more accurate at determining the valence of the conveyed emotion when the haptic sensation was displayed first. However, this is the opposite of the results from Phase II, where participants were more accurate at determining valence when the audio was displayed first. It is likely that there is an order effect causing this shift since participants had the chance to become more comfortable using the system during Phase I. They also learned a little bit about each other and the general auditory emotion language of their partner. Another possible reason for this shift was the change in prompts between the two experimental phases. In Phase I, participants were asked to convey discrete emotions, whereas in Phase II they were given broader categories of valence. Therefore, participants had more choice in the emotion they chose to convey. Overall, participants found it more straightforward to distinguish discrete emotions when displayed as touch and broader categories of valence when conveyed as audio.

Our analysis of the audio features suggests that the tonal chord features (that can be directly associated with pitch and notes) have a high significance for a given emotional prompt. Since the haptic interface uses notes and pitch values from the audio interface to determine the haptic mapping, it is safe to suggest that features used by the haptic interface do carry emotional significance. Further, results from our demographic and contextual analysis looking at spectral features, country of origin, and closeness of users, points to correlations between closeness and felt emotion, and significant variation among vocalized spectral features depending on region (North America versus Asia).

B. Suggested Study

An important next step would be to study the crossmodal affective communication of our system in the field, in different user situations.This means conducting contextual user studies

that explore situations where users can come up with their own vocal expressions without receiving emotional prompts. Since our system does not assume a stabilized lexicon or consideration of affect, the personalizability of our system requires us to study the vocal to haptic affective communication potentials of our design in contexts closer to users' daily scenarios.

C. User Context and Relationships in Interactions

The need to consider non-Western scales is supported by data analysis insights showing that there is a correlation between specific vocalization parameters such as spectral flux, and user country of origin, supporting the need for context sensitivity in creating affective systems that facilitate emotional communication between users. We found differences in vocalization specifically with spectral flux. The spectral flux was significantly higher for someone from North America than when the vocalizer was from Asia.

Of our 20 individuals comprising 10 user pairs. Some pairs were from the same country of origin (and actually sang to each other in non-English language) and some were from different countries of origin. There were only 4 pairs of users from the same country of origin. One pair identified as both being from China and three pairs self-identified as being from the USA. The rest of the 6 pairs were not from the same country of origin.

A concern we raise is our choice of using the Western scale. Amongst our users, not all were native English speakers, and during our user study, not all chose to vocalize (hum or sing) songs from Western musical scales. We cannot assume that a Western scale is suited to all affective states, and we cannot require all users from non-Western backgrounds to map their expression onto a system that uses a Western musical model. Instead, it is important to consider how our design might support and include non-Western musical preferences and imaginations. We can look to the field of postcolonial computing which explores the ways in which colonialism and imperialism have shaped the development and deployment of technology and computing [37] [38]. This is one framework that can support haptic designers in considering what equitable and culturally responsive approaches to socially affective systems development and haptic design might be. The next step in our project is to consider how to design for different (non-Western) scales. Non-Western musical scales refer to musical scales used in musical traditions outside of the Western classical music tradition and often have a distinct set of notes, intervals, and tuning systems. Two considerations are microtonality and the modal system. Microtonality refers to the use of finer division of pitch intervals. For example, Indian classical music uses a 22-note system of microtones [39]. Secondly, modality refers to differing modal systems, which may use a limited number of notes and include repetition of specific musical patterns. As an example, Arabic music uses a system of "maqam" modes, which consist of specific intervals and melodic patterns [40], [41]. These are important tonal and pattern differences to consider when designing a system that captures and transposes musical expression. With respect to our design, this means adjusting how we map frequencies detected, and also how we design the hardware and actuators to reflect the notes of non-Western scales.

Designing for user-to-user customization can increase user satisfaction and engagement [42]. However, designing for user-to-user customization requires making sure that technological systems are accessible and usable for all, including users with disabilities [43]. Users with disabilities benefit from customized technology. This includes designing systems that give users agency to change them to suit their needs as exemplified by the "Do It Yourself Assistive Technology (DIY-AT)" design approach [44]. With this in mind, the next step in our prototyping stage is to expand to considering users with motor, sensory and cognitive disabilities.

VI. Conclusion

This paper considers how paired users may convey idiosyncratic affective communication. Our design considers what it means to create affective systems which do not assume the need to stabilize or require a (system-mediated) matching of emotional states between users to be effective. Considering that the effectiveness of the socially affective design is influenced by individual differences in personality and cultural background [45], our approach of designing and testing a system that allows for idiosyncratic expression that is not matched or corrected by the system leads to more user agency and more inclusive affective experience.

Ethical Impact Statement

Our ethical issues mainly concern human subject research. This research was approved by the University of Southern California's Institutional Review Board (IRB) under Protocol #UP-20-01131. All participants were adults and gave informed consent. The identities of our 20 users were considered, further, we considered ways that cultural context and interpersonal relationships intersect with affective research. When considering the demographics of our users, we acknowledge the difficulty (and geopolitical tensions) in categorizing geographic areas (when looking at results and grouping users by country of origin).

We include examples (references) of ethical critiques of affective systems, including critiques of ways that user systems and affective computing systems can be designed to exclude users and demographics.

We discuss approaches to engage in inclusive design approaches. This includes addressing cross-cultural impacts in affective design and user testing of affective systems. This includes the ethical concern of defaulting a musical system to being designed for the Western scale, and how this may limit affective imagination and expression for users from other musical traditions and geographies. This ethical concern also intersects with the use of machine learning models. Our model confronts a cultural bias toward the Western musical scale which we critique in our paper. We explore ways to design for non-Western musical scales.

REFERENCES

[1] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, and A. R. Jaskolka, "Touch communicates distinct emotions." *Emotion*, vol. 6, no. 3, p. 528, 2006.

[2] S. M. Erk, A. Toet, and J. B. Van Erp, "Effects of mediated social touch on affective experiences and trust," *PeerJ*, vol. 3, p. e1297, 2015.

[3] D. Tsetserukou, "Haptihug: A novel haptic display for communication of hug over a distance," in *Haptics: Generating and Perceiving Tangible Sensations: International Conference, EuroHaptics 2010, Amsterdam, July 8-10, 2010. Proceedings, Part I*. Springer Berlin Heidelberg, 2010, pp. 340–347.

[4] a. G. Barrett, Mesquita, "Context in emotion perception," *Current Directions in Psychological Science*, vol. 19(5), p. 317–322, 2010.

[5] V. J. Harjunen, M. Spapé, I. Ahmed, G. Jacucci, and N. Ravaja, "Individual differences in affective touch: Behavioral inhibition and gender define how an interpersonal touch is perceived," *Personality and Individual Differences*, vol. 107, pp. 88–95, 2017.

[6] M. Salvato, S. R. Williams, C. M. Nunez, X. Zhu, A. Israr, F. Lau, K. Klumb, F. Abnousi, A. M. Okamura, and H. Culbertson, "Data-driven sparse skin stimulation can convey social touch information to humans," *IEEE Transactions on Haptics*, vol. 15, no. 2, pp. 392–404, 2022.

[7] K. R. Scherer, "Vocal affect expression: a review and a model for future research." *Psychological bulletin*, vol. 99, no. 2, p. 143, 1986.

[8] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *J. of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[9] J. Gratch, "The field of affective computing: An interdisciplinary perspective," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 31, no. 1, 2021.

[10] K. Crawford, R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, A. Kak, V. Mathur, E. McElroy, A. N. Sánchez *et al.*, "Ai now 2019 report," *New York, NY: AI Now Institute*, 2019.

[11] R. W. Picard, *Affective computing*. MIT press, 2000.

[12] S. Holland, A. J. Bouwer, M. Dalgelish, and T. M. Hurtig, "Feeling the beat where it counts: fostering multi-limb rhythm skills with the haptic drum kit," in *Proc. Conference on Tangible, embedded, and embodied interaction*, 2010, pp. 21–28.

[13] E. Partesotti, A. Peñalba, and J. Manzolli, "Digital instruments and their uses in music therapy," *Nordic Journal of Music Therapy*, vol. 27, no. 5, pp. 399–418, 2018.

[14] B. Dörr, F. Norouzinia, K. Altmeyer, and D. Werth, "Haptic technology in digital music learning context: A state-of-the-art analysis," in *European Conference on e-Learning*, vol. 21, no. 1. Academic Conferences International Limited, 2022, pp. 87–94.

[15] I. Hwang, H. Son, and J. R. Kim, "Airpiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback," in *Proc. IEEE World Haptics Conference*, 2017, pp. 213–218.

[16] H.-J. Suk, S.-H. Jeong, T.-H. Yang, and D.-S. Kwon, "Tactile sensation as emotion elicitor," *Kansei Engineering International*, vol. 8, no. 2, pp. 153–158, 2009.

[17] X. Zhu, T. Feng, and H. Culbertson, "Understanding the effect of speed on human emotion perception in mediated social touch using voice coil actuators," *Frontiers in Computer Science*, vol. 4, 2022. [Online]. Available: https://par.nsf.gov/biblio/10392584

[18] J. B. Van Erp and A. Toet, "Social touch in human–computer interaction," *Frontiers in digital humanities*, vol. 2, p. 2, 2015.

[19] R. Wang, F. Quek, D. Tatar, K. S. Teh, and A. Cheok, "Keep in touch: channel, expectation and experience," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 139–148.

[20] J. Kjeldskov, M. Gibbs, F. Vetere, S. Howard, S. Pedell, K. Mecoles, and M. Bunyan, "Using cultural probes to explore mediated intimacy," *Australasian Journal of Information Systems*, 2005.

[21] H. Seifi and K. E. Maclean, "A first look at individuals' affective ratings of vibrations," in *Proc. World Haptics Conference*, 2013, pp. 605–610.

[22] J. Stets and J. e. Turner, "Cultural theory and emotions," *Handbook of the Sociology of Emotions*, 2006.

[23] D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations," *Proceedings of the National Academy of Sciences*, vol. 107(6), p. 2408–2412, 2010.

[24] S. C. Aker, K. F. Faulkner, H. Innes-Brown, M. Vatti, and J. Marozeau, "Audio-tactile congruences in vibro-tactile music enhancement world haptics conference," in *Proc. IEEE World Haptics Conference*. IEEE, 2021, pp. 345–345.

[25] R. Paisa and A. Bargum, "Team musical armband," in *IEEE World Haptics Conference Student Innovation Challenge*, 2021. [Online]. Available: https://github.com/WHC2021SIC/WHC2021SIC-TeamMusicalArmband/blob/master/README.md

[26] M. B. Mokhsin, N. B. Rosli, W. A. W. Adnan, and N. A. Manaf, "Automatic music emotion classification using artificial neural network based on vocal and instrumental sound timbres." in *SoMeT*, 2014.

[27] A. Tursunov, S. Kwon, and H.-S. Pang, "Discriminating emotions in the valence dimension from speech using timbre features," *Applied Sciences*, vol. 9, no. 12, p. 2470, 2019.

[28] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proc. ISMIR*, vol. 86, 2010, pp. 937–952.

[29] M. Mauch and S. Dixon, "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 659–663.

[30] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[31] e. gomez and e. maestre, "automatic characterization of dynamics and articulation of expressive monophone recordings," *journal of the audio engineering society*, may 2005.

[32] A. Webb and J. Peck, "Individual differences in interpersonal touch: On the development, validation, and use of the "comfort with interpersonal touch"(cit) scale," *Journal of Consumer Psychology*, vol. 25, no. 1, pp. 60–77, 2015.

[33] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness." *Journal of personality and social psychology*, vol. 63, no. 4, p. 596, 1992.

[34] J. Brotzer, E. Mosqueda, and K. Gorro, "Predicting emotion in music through audio pattern analysis," in *IOP Conference Series: Materials Science and Engineering*, vol. 482, no. 1, 2019, p. 012021.

[35] P. Herrera and S. Streich, "Detrended fluctuation analysis of music signals: Danceability estimation and further semantic characterization," in *Audio Engineering Society Convention 118*, 2005.

[36] R. Panda, R. M. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: a survey," *IEEE Transactions on Affective Computing*, 2020.

[37] S. D. Mainwaring and P. Dourish, "Ubicomp's colonial impulse," in *Proc. ACM Conference on Ubiquitous Computing*, 2012, pp. 133–142.

[38] L. Irani, J. Vertesi, P. Dourish, K. Philip, and R. E. Grinter, "Postcolonial computing: a lens on design and development," in *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2010, pp. 1311–1320.

[39] N. A. Jairazbhoy and A. Stone, "Intonation in present-day north indian classical music," *Bulletin of the School of Oriental and African Studies*, vol. 26, no. 1, pp. 119–132, 1963.

[40] S. Yöre, "Maqam in music as a concept, scale and phenomenon," *Zeitschrift für die Welt der Türken/Journal of World of Turks*, vol. 4, no. 3, pp. 267–286, 2012.

[41] B. Nettl, "Music of the middle east," *Excursions in world music*, pp. 46–73, 2001.

[42] J. S. Brown, A. Collins, and P. Duguid, "Situated cognition and the culture of learning," *Subject Learning in the Primary Curriculum: Issues in English, Science*, vol. 18, no. 1, pp. 32–42, 1989.

[43] B. Shneiderman, C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.

[44] J. Hook, S. Verbaan, A. Durrant, P. Olivier, and P. Wright, "A study of the challenges related to diy assistive technology in the context of children with disabilities," in *Proc. Conference on Designing Interactive Systems*, 2014, pp. 597–606.

[45] R. A. Calvo and D. Peters, *Positive computing: technology for wellbeing and human potential*. MIT press, 2014.