Music Tactalizer: A Wearable Haptic Music Player with Multi-Feature Audio-Tactile Rendering

Aaron Raymond See*
Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung City, Taiwan
aaronsee@gm.ncut.edu.tw

Aaron Benjmin R. Alcuitas
Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung City, Taiwan
aaronalcuitas16@gmail.com

Thad Jacob T. Tiong
Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung City, Taiwan
thadtiong00@gmail.com

Vence Jumar A. Sasing
Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung City, Taiwan
rjsasing@gmail.com

2025 Seventh International Symposium on Computer, Consumer and Control (IS3C) | 979-8-3315-8700-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/IS3C65361.2025.11131083

Abstract— **Haptic music player wearables use vibrotactile feedback to improve the immersion and quality of the music listening experience. The most common rendering approach is transducing the audio into tactile signals, but the human tactile perception capacity requires a frequency limit of 1 kHz, heavily limiting the communication of music features. In this study, we introduce the Music Tactalizer, a wearable haptic music player that enriches the music listening experience by embedding pitch, rhythm, melody, and timbre information through audio-tactile mapping. The algorithm extracts the music features and converts them to tactile signals using digital signal processing. Through a mobile app that provides synchronized audio and visual feedback, the tactile signals are transmitted to a wearable using Bluetooth Low Energy to deliver vibrotactile feedback to the user's fingertips. The generated tactile signals were validated using mean average error (MAE), dynamic time warping (DTW) distance, and accuracy, yielding normalized averages of 0.1240 MAE and 0.1392 DTW distance, and a 78.24% rhythmic pattern accuracy. The Music Tactalizer can communicate multiple musical features at an auditory frequency range while retaining most of the original audio signal's characteristics. Future research can improve the quantitative results or explore other methods and applications.**

Keywords—haptic music player, wearable device, audio-tactile rendering, digital signal processing

I. Introduction

Vibrotactile feedback has been found to enhance the immersive and overall quality of music [1, 2]. In particular, vibration signals with dynamics similar to the corresponding audio provided more enjoyment compared to non-vibrating and noisy vibrating setups [1], and vibrations in time with music were used in virtual reality (VR) concerts to recreate feelings of psychological connection and presence in live concerts [2]. Employing the positive effects of vibrotactile feedback towards the music listening experience, multiple studies have been dedicated to allowing both the able and hearing impaired to "feel" music through haptic music players (HMPs) with audio-tactile rendering. HMPs range from fixed setups or installations, to wearable devices (HMP-WD), to hybrid setups, with most of the research attention understandably shifting towards HMP-WDs [3].

HMP-WDs have emerged in various forms such as belts, bracelets, and suits, but have gradually shifted to a glove form factor [4-8]. The decision on form is based on the high concentration of tactile receptors in the hands, making it most sensitive to vibrations [4-6]. The most commonly used audio-tactile rendering approach by HMP-WDs is transducing the audio to tactile signals, with little to no additional signal processing, so that the skin feels vibrations similar to what the ear hears. Voice coil actuators are often used for this purpose [4, 5, 7]. Some gloves also use coin motors [6, 8], but since the frequency of vibrations cannot be controlled in this setup, the vibrations are instead matched to the beat of the song. These studies have successfully established the feasibility and effectivity of mapping audio to tactile signals that enhance the music listening experience.

The weakness of prior studies, however, is that the rendered frequency range is heavily constrained. Even if the motors used have higher capabilities, the human tactile perception range is only 5-1000 Hz, while it is 20-20000 Hz for audio [4, 6]. This indicates that much audio information is never communicated in current audio-tactile rendering methods, as prior HMP-WDs limit their frequencies to 1000 Hz. Furthermore, this allows only a few musical features to be displayed at a time, such as pitch, rhythm, or tempo, but rarely a combination of these features. Alternative approaches such as the use of tactile illusions have been explored, but these are subject to the same constraints and gaps [9]. Table I presents the current state of research on audio-tactile rendering from a recent review [3]. Expanding the communicated frequency range and music features requires exploring other actuators and newer methods so that HMP-WDs can finally display frequencies that can be heard but not felt.

TABLE I. Overview of Research on HMP-WD Gloves

HMP-WD Glove	Actuators	Stimuli	Features	Frequency Limit (Hz)
Audio-Tactile Glove [4]	VC[a]	VA[d]	Pitch	1000*
Mood Glove [5]	VC	VAV[e]	Rhythm, VMP[g]	1000*
Vibrotactile Captioning [6]	ERM[b]	VV[f]	Tempo, VMP	1000*
SENSE [7]	VC	VV	VMP	1000*
Haptic gloves to elicit emotions [8]	ERM	VA	Tempo	1000*
Music Tactalizer	P[c]	VAV	Pitch, Rhythm, Melody, Timbre	12000

[a.] Voice coil, [b.] Eccentric Rotating Mass, [c.] Piezoelectric, [d.] Vibrotactile-Auditory, [e.] Vibrotactile-Auditory-Visual, [f.] Vibrotactile-Visual, [g.] Vibrotactile music or perception with or without music listening, * Due to only accounting for cutaneous-detectable frequencies

This work was supported in part by the National Science and Technology Council of Taiwan under grant numbers NSTC 113-2221-E-167-049 and the Ministry of Education of Taiwan under grant number 13001110179-EDU.

In this study, we introduce the Music Tactalizer, an HMP-WD that enriches the music listening experience by embedding pitch, rhythm, melody, and timbre information through vibrotactile signals. The system is composed of haptic gloves equipped with piezoelectric actuators that provide vibrotactile feedback, managed by a mobile app that synchronizes the music with the tactile signals. The unique capability of piezoelectric actuators for localization, as exhibited in prior research [10], is employed to display the separation of multiple instruments, the changes of pitch over time, and the rhythmic patterns produced by music. This work may provide the means to break current limitations on communicable frequency ranges in audio-tactile rendering.

II. Materials and Methods

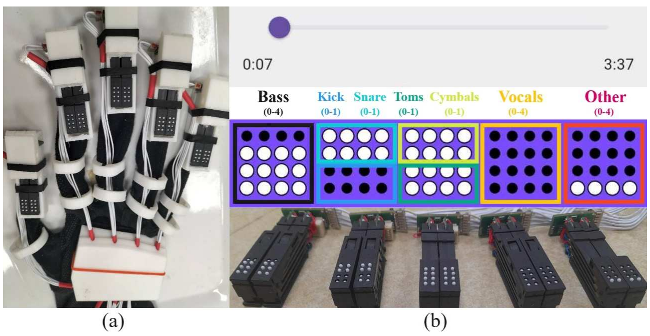
The Music Tactalizer has three main components: the haptic glove, the mobile app, and the rendering system. The rendering system takes an audio file and creates tactile patterns using the algorithm that will be discussed further in this paper. The tactile patterns are sent to the mobile app, which sends them to the haptic glove for vibrotactile feedback synchronized with the app's visual and audio feedback. The haptic glove receives the patterns and activates the pins accordingly. Figure 1 presents the Music Tactalizer system.

A. Haptic Gloves

The haptic gloves were developed using off-the-shelf gloves equipped with P20 braille cells and a custom printed circuit board (PCB) with an STM32F10 microcontroller chip, all held together by 3D printed cases for housing. The PCB is equipped with a JDY33 Bluetooth module to enable a Bluetooth Low Energy (BLE) interface in slave mode, allowing it to receive tactile patterns from the software. The braille cells deliver a minimum tactile force of 17 cN and a rapid dot rising time of 50 ms, both of which fall within the human detection threshold of 10 cN and 1-100 ms. Each braille cell consists of eight dots, and two braille cells are used for each finger. The methods employed by the haptic gloves have shown success in prior studies for providing localized cutaneous force feedback in robotics assisted surgery systems [10] and virtual reality immersion [11], with its capacity for localization showing potential for HMP applications.

B. Mobile Application

The mobile app presents visual and audio stimuli synchronized with the haptic gloves' vibrotactile stimuli. The app connects to the haptic gloves via BLE for seamless communication, enabling the app to send tactile patterns as the music plays in real time. The tactile patterns assign music elements to different fingers, such that the percussive elements are closer to the index finger, which is commonly observed in finger-tapping studies [12]. Founded on this principle, the kick and snare are assigned to the index finger and the toms and cymbals are assigned to the middle finger. While the percussive elements observe no pitch, they are ordered from lowest to highest in terms of frequency range and serve to display the rhythmic patterns of a song. The melodic elements that have pitch are assigned to the remaining fingers. The bass, which has percussive properties and is the lowest-pitched, is assigned to the thumb. The vocals are assigned to the ring finger and others occupy the pinky finger. The rows of the pin grid communicate pitch information on four levels, with the activation of only the bottom row showing the lowest pitch while the activation of all rows showing the highest pitch. In this setup, the separation of instruments communicate timbre while the movements of pitch simulated in the vertical movement of pins communicate melody.



(a) (b)

Fig. 1. The Music Tactalizer: (a) the haptic gloves equipped with the vibrotactile elements, and (b) the mobile app and feature-pin mapping.

C. Audio-Tactile Rendering Algorithm

The rendering system extracts music features from an audio file and converts them to tactile signals. The algorithm consists of three steps: (1) source separation, (2) pitch tracking and peak detection, and (3) discretization and mapping. Figure 2 illustrates the audio-tactile rendering algorithm.

The first step involves separating the source audio file into various instrument audio files. We used a pre-trained hybrid separation model [13] that applies spectrogram and waveform separation using Transformers to separate the audio to four instrument files: drums, bass, vocals, and other. Since the drum is an unpitched percussion, the drums audio file is further separated into four parts: kick, snare, toms, and cymbals. The respective frequency splits were heuristically determined based on where each drum is best captured, which became 30, 150, 800, 5000, and 12000 Hz. Given these splits, a Butterworth bandpass filter was applied to the drums audio to create four more audio files for each of the subsets.

The second step involves extracting the pitch of the melodic audio and the peaks of the percussive audio. For the melodic audio, a short-time Fourier transform (STFT) is applied to attain a time-frequency matrix. The matrix is used in estimating the pitches and their magnitudes through quadratically interpolated fast Fourier transform, an approximate maximum likelihood method for spectral peak estimation [14]. The pitch detections are cleaned using thresholding. For each time frame, the dominant pitch is extracted and outlier pitch values are removed. For the percussive audio, STFT is also applied and local maxima that represent peaks in the energy signal, which estimate the points where the percussion instrument was used, are identified.

The third step involves converting the continuous pitch contours to four discrete values based on the quartile they belong to relative to the maximum pitch value in the audio file, while time points with no pitch values are set to zero. Both the discretized pitch contours and the energy peaks are assigned to 100-ms time bins so that the tactile patterns are uniformly spaced. The energy peaks are converted to binary, determined by the existence of a peak in the time bin. With the discretized pitch and peak values, we process the results for all elements to a structured dictionary and then to strings representing the tactile patterns that would be sent to the hardware. The tactile patterns generated in this final step correspond to the feature-pin mappings, completing the audio-tactile rendering process.

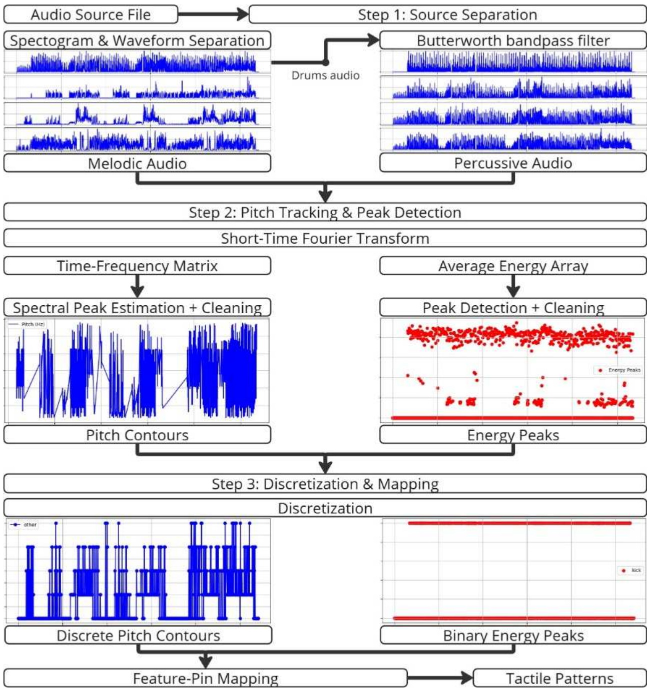


Fig. 2. The multi-feature audio-tactile rendering algorithm used by the Music Tactalizer.

III. Results and Discussion

To prove that the Music Tactalizer could display the musical features pitch, melody, timbre, and rhythm while retaining a significant portion of the original audio's melodic contour and rhythmic patterns at its complete frequency range, we perform signal fidelity testing with respect to each feature. For melodic elements, the pitch display is evaluated using mean absolute error (MAE) to quantify the deviation of the tactile pitch patterns from the original, cleaned pitch contours. Displays of melody, which is the movement of pitch over time, and timbre, which is the distinctive quality observed between instruments independent of pitch, are evaluated by calculating their Dynamic Time Warping (DTW) distance from the same pitch contours. For percussive elements, the rhythm display is evaluated using accuracy, which quantifies how much of the tactile actuations corresponded to the energy peaks representing beats from the original audio. In this experiment, the mobile app recorded the tactile signals sent to the haptic glove and their time points in a log file. The five songs were chosen arbitrarily, but we ensured that all had vocals and were from different genres, as has been done by similar studies [15]. The log file was used to recreate the tactile signals, which would be compared with the audio signals. Table II summarizes the testing results.

TABLE II. Signal Fidelity Testing Results

Metric		**Average**	
		Over 5 songs	***Overall***
MAE[a]	Bass	0.1151	0.1240
	Vocals	0.1296	
	Other	0.1272	
DTW distance[a]	Bass	0.1015	0.1392
	Vocals	0.1931	
	Other	0.1230	
Accuracy	Kick	0.8163	0.7824
	Snare	0.7867	
	Toms	0.7633	
	Cymbals	0.7633	

[a] In normalized values for pitch (Hz)

A. Pitch Display

The pitch levels of the tactile signals were compared with those of the cleaned original audio. For each stem, the pitch values were normalized, then MAE was computed using:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{1}$$

where N is the number of time bins, y is the normalized pitch from the audio, and $\hat{y}$ is the normalized pitch from the tactile signals. The MAE reveals how far apart the two pitch contours are on average, and a lower MAE means the tactile signals communicate the pitch of the original audio more accurately. The average MAE for all melodic elements is 0.1240. Figure 3 visualizes a comparison between the normalized tactile and audio signals. Due to discretization, the tactile pitch display deviates from the original audio but manages to closely resemble the heights and depths of the pitch contour.

B. Melody/Timbre Display

DTW is used to measure the similarity of the audio and tactile pitch sequences independent from time alignment. The Euclidean distance is calculated between points from both sequences to construct a distance matrix, which is used to recursively compute the accumulated cost matrix. The results are used to extract the warping path. The normalized DTW distance is then calculated using:

$$DTW = \frac{1}{K}\sum_{i=1}^{K} d\left(a_{i_k}, b_{j_k}\right) \tag{2}$$

where K is the length of the warping path, (a, b) is the DTW-aligned pair at index k, where a is a value from the audio signal while b is a value from the tactile signal, and their respective indices i and j, and d is the Euclidean distance between the two values. The average normalized DTW distance for all melodic elements is 0.1392. Figure 4 presents examples of the DTW alignment path between the audio and tactile pitch sequences.

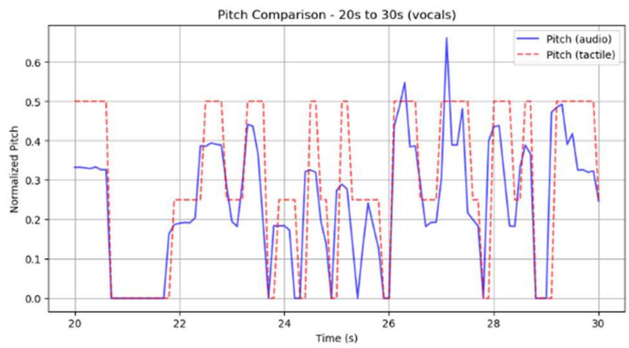


Fig. 3. Pitch contour comparison plot of a song's vocal stem.

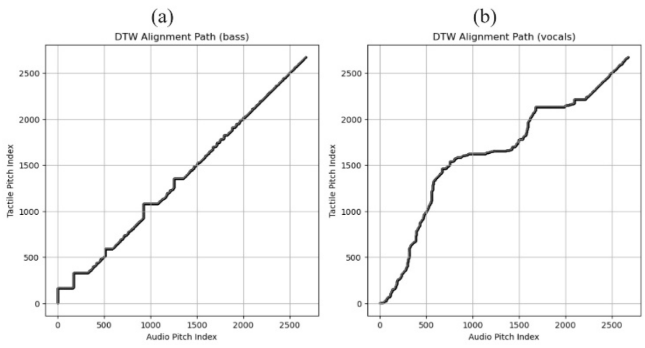


Fig. 4. DTW alignment paths for a song: (a) bass and (b) vocals.

In this example, the DTW path is a near-perfect diagonal line for the bass stem, as most melodic contours are aligned with little warping needed. However, the DTW path for the vocals bend from the diagonal, showing temporal misalignment due to the latency and lower resolution of the tactile signal.

C. Rhythm Display

Accuracy is used to evaluate the closeness of the rhythm of the tactile signal to that of the original audio. For each 100-ms time bin, it is ideal that the pins actuate if and only if an energy peak is observed in the original audio, otherwise, it does not fully communicate the rhythm. The accuracy is calculated using:

$$ACC = \frac{1}{N}\sum_{i=1}^{N} 1(y_i = \hat{y}_i) \quad (3)$$

where N is the number of time bin, y is the normalized pitch from the audio, $\hat{y}$ is the normalized pitch from the tactile signals, and $1(y=\hat{y})$ is an indicator function that returns 1 when the tactile and audio signals match and 0 otherwise. Figure 5 shows an example of tactile pin actuations compared to the beats from the original audio. In most cases, the tactile actuations match the audio's beats, but latency causes instances where the pins reset late. As a result, some beats from the audio are not displayed. The average accuracy for all percussive elements is 78.24%.

IV. Conclusion

The Music Tactalizer's tactile display of pitch, melody, timbre, and rhythm attained sufficiently high scores from their respective metrics, considering the portability and efficiency of methods used. The pitch display has an average MAE of 0.1240, which is relatively low but not negligible, and could be reduced by increasing the number of pitch levels during discretization. The melody/timbre display has an average normalized DTW distance of 0.1392, showing that the tactile pitch contour generally follows the shape of the audio signals. Increasing the number of pitch levels could also improve this by displaying finer movements of pitch. The rhythm display demonstrated an accuracy of 78.24%, which shows that the device replicates the majority of the beats from the audio.

This study introduces a novel audio-tactile rendering method for HMP-WD systems, addressing prior frequency limitations by enabling tactile translation of audio features beyond 1000 Hz. It also proposes validation techniques for rendering of musical features, which were left unverified in previous studies [6]. The device employs piezoelectric actuators for precise point localization and supports portability via BLE, while prior studies seldom explored the use of Bluetooth [7]. Consistent with glove-based HMP-WD studies, this work targets fingertip stimulation [6-8], unlike prior efforts focused on the phalanges, backhand, or palm [4, 5]. While human tactile sensitivity peaks at 100–400 Hz [4], this study mitigates frequency response variability through a frequency-mapping strategy. As in related works [5, 6, 8], vibrotactile stimuli can also potentially evoke emotional responses akin to auditory cues.

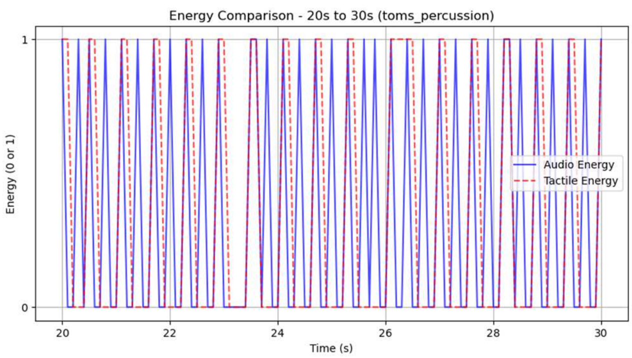


Fig. 5. Comparison of energy actuations between audio and tactile signals of the "toms" frequency subset from the drums stem of a song.

The Music Tactalizer represents an advancement in HMP-WD technology, maintaining signal fidelity while displaying multiple music features through audio-tactile rendering. Future work will focus on improving the system, expanding its applications, and facilitating usability testing to validate the subjective music experiences evoked by the device.

References

[1] S. Papetti, H. Järveläinen, and S. Schiesser, "Interactive vibrotactile feedback enhances the perceived quality of a surface for musical expression and the playing experience," *IEEE Transactions on Haptics,* vol. 14, no. 3, pp. 635-645, 2021.

[2] T. Venkatesan and Q. J. Wang, "Feeling connected: The role of haptic feedback in VR concerts and the impact of haptic music players on the music listening experience," in *Arts*, 2023, vol. 12, no. 4: MDPI, p. 148.

[3] B. Remache-Vinueza, A. Trujillo-León, M. Zapata, F. Sarmiento-Ortiz, and F. Vidal-Verdú, "Audio-tactile rendering: a review on technology and methods to convey musical information through the sense of touch," *Sensors,* vol. 21, no. 19, p. 6575, 2021.

[4] G. Young, D. Murphy, and J. Weeter, "Audio-tactile glove," in *Proceedings of the International Conference on Digital Audio Effects (DAFx), Maynooth, Ireland*, 2013.

[5] A. Mazzoni and N. Bryan-Kinns, "Mood glove: A haptic wearable prototype system to enhance mood music in film," *Entertainment Computing,* vol. 17, pp. 9-17, 2016.

[6] M. J. Lucía *et al.*, "Vibrotactile captioning of musical effects in audio-visual media as an alternative for deaf and hard of hearing people: An EEG study," *IEEE Access,* vol. 8, pp. 190873-190881, 2020.

[7] K. Enriquez, M. Palacios, D. Pallo, and G. Guerrero, "SENSE: Sensory component VR application for hearing impaired people to enhance the music experience," in *2020 15th Iberian conference on information systems and technologies (CISTI)*, 2020: IEEE, pp. 1-6.

[8] G.-L. Alvaro, V. B. Ricardo, S. P. J. Manuel, O. Tomás, and C. Víctor, "New haptic systems for elicit emotions in audio-visual events for hearing impaired people," *Procedia Computer Science,* vol. 237, pp. 533-543, 2024.

[9] B. Remache-Vinueza *et al.*, "Mapping monophonic MIDI tracks to vibrotactile stimuli using tactile illusions," in *International Workshop on Haptic and Audio Interaction Design*, 2022: Springer, pp. 115-124.

[10] A. R. See, T. J. Tiong, L. B. De Guzman, K. Contee, G. G. Lim, and C. S. Yebes, "Development of Localized Cutaneous Force Feedback System for Robotics Assisted Surgery Systems," *Procedia Computer Science,* vol. 246, pp. 1160-1169, 2024.

[11] A. C. Abad, D. Reid, and A. Ranasinghe, "A novel untethered hand wearable with fine-grained cutaneous haptic feedback," *Sensors,* vol. 22, no. 5, p. 1924, 2022.

[12] Q. Sun, S. Li, Z. Yao, Y.-L. Feng, and H. Mi, "PalmBeat: A Kinesthetic Way to Feel Groove With Music," in *12th augmented human international conference*, 2021, pp. 1-8.

[13] A. Défossez, N. Usunier, L. Bottou, and F. Bach, "Demucs: Deep extractor for music sources with extra unlabeled data remixed," *arXiv preprint arXiv:1909.01174,* 2019.

[14] J. O. Smith III, "Spectral audio signal processing," *(No Title),* 2011.

[15] F. Alves Araujo, F. Lima Brasil, A. Candido Lima Santos, L. de Sousa Batista Junior, S. Pereira Fonseca Dutra, and C. Eduardo Coelho Freire Batista, "Auris system: Providing vibrotactile feedback for hearing impaired population," *BioMed research international,* vol. 2017, no. 1, p. 2181380, 2017.