Gesture-Driven DDSP Synthesis for Digitizing the Chinese Erhu

Wenqi WU
Computational Media and Arts
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
wwu252@connect.hkust-gz.edu.cn

Hanyu QU
Computational Media and Arts
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
hqu817@connect.hkust-gz.edu.cn

ABSTRACT

This paper presents a gesture-controlled digital Erhu system that merges traditional Chinese instrumental techniques with contemporary machine learning and interactive technologies. By leveraging the Erhu's expressive techniques, we develop a dual-hand spatial interaction framework using real-time gesture tracking. Hand movement data is mapped to sound synthesis parameters to control pitch, timbre, and dynamics, while a differentiable digital signal processing (DDSP) model, trained on a custom Erhu dataset, transforms basic waveforms into authentic timbre which remians sincere to the instrument's nuanced articulations. The system bridges traditional musical aesthetics with digital interactivity, emulating Erhu bowing dynamics and expressive techniques through embodied interaction. The study contributes a novel framework for digitizing Erhu performance practices, explores methods to align culturally informed gestures with DDSP-based synthesis, and offers insights into preserving traditional instruments within digital music interfaces.

Author Keywords

DDSP, Erhu, Gesture, Chinese Instrument, Interactive Music Performance

1. INTRODUCTION

The rapid breakthroughs in sensor fusion[22, 2, 18, 10] and artificial intelligence[14, 13, 9] in recent decades have led to significant improvements in both accuracy and real-time responsiveness. Researchers and artists have widely applied these technologies to various domains. These include the development of Digital Musical Instruments (DMIs)[20, 3], augmented instruments[7, 16, 21], and interactive music performance[19, 1, 17]. These advancements offer multiple innovative possibilities for contemporary music creation and performance.

As one of the oldest and most influential traditional instruments in China, the Erhu has played a pivotal role in Chinese and East Asian musical culture [6, 8]. Despite having only two strings, it possesses a remarkably rich and nuanced expressiveness. The Erhu' s cultural value is evident not only in its classic repertoire but also in its universal appeal that spans regions, ethnicities, and eras[6]. However, while the global community increasingly gravitates toward Western instruments and culture during the process of globalization[4], the Erhu and other traditional Chinese instruments have received comparatively limited attention in both technological research and artistic practice[4].

Drawing on traditional Erhu performance techniques, we develop a gesture-controlled air instrument reminiscent of the Theremin paradigm. By leveraging spatial hand gesture tracking, our system enables real-time mapping of embodied gestures to core sound synthesis parameters—specifically, pitch, timbre, and dynamic expression in the Max[1] environment. Unlike traditional Erhu performance, which relies on physical string manipulation and bowing, this digital interface abstracts these techniques into intuitive spatial interactions. For instance, vertical hand movements control pitch (analogous to Theremin's pitch antenna), while horizontal movements modulate volume (mirroring Theremin's volume antenna). Concurrently, we integrate a DDSP (Differentiable Digital Signal Processing)[5] model trained on authentic Erhu recordings to transform raw waveforms into a timbre that preserves the instrument's nuanced articulations (e.g., vibrato, glissando).

Throughout the machine learning training and interaction design process, we place special emphasis on preserving and exploring the Erhu' s acoustic characteristics and classical techniques, while incorporating reflections on traditional performance experiences into the overall DMI design. Our project offers the following contributions: (1) creating novel interactive music performance contexts for the Erhu as a traditional Chinese instrument; (2) introducing new possibilities and inspirations for digital instruments and air instruments by drawing on the performance techniques of the Erhu and other traditional Chinese instruments; and (3) providing the Chinese NIME community with a practical example that integrates traditional instruments and digital innovation.

We believe that the development of this Erhu performance system will further expand the instrument' s applications in electronic music and interactive art, serving as a more flexible and creative bridge between tradition and modernity, as well as among different regions and cultures. We hope that, in future research and practice, this project may offer new ideas and references for the digital and global evolution of other Chinese traditional instruments.

[1] https://cycling74.com/

Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'25). June 24–27, 2025. The Australian National University, Canberra, Australia.

2. BACKGROUND

2.1 Cultural Background of Erhu

The Erhu is a two-stringed bowed musical instrument from China with a long history, originating from the bowed-string tradition of the Tang and Song dynasties. Its design is believed to have evolved from the "yazheng" or "xiqin," initially serving as an accompaniment instrument for operas and folk singing before gradually being integrated into various ensemble and solo contexts[6]. Over the past few centuries, the Huqin family (a family of spike fiddle popularly used in traditional Chinese music)[2] has expanded significantly, giving rise to numerous variants in addition to the Erhu, such as the Banhu, Gaohu, Jinghu, and the Morin Khuur (Horsehead Fiddle). In the early twentieth century, the Erhu began its professionalization, fueled by classic repertoire composed by musicians including Liu Tianhua[12] and Hua Yanjun (Abing)[15], thereby opening new avenues for modern Chinese instrumental performance[6]. With its distinctive national character and vocal-like timbre, the Erhu now plays a vital role in both traditional and contemporary music settings.

At the core of Erhu performance lies the coordination between the right hand's bowing and the left hand's fingering. The right hand shapes the timbre through various combinations of bow speed, bow pressure, and bowing segments —such as long bow, short bow, accented bow, and spiccato-like techniques; in modern works, extended approaches like tapping the strings with the bow stick or reversing the bow hair have also emerged (Liu Changfu, 2018)[11]. Meanwhile, the left hand focuses on capturing the instrument' s "vocal resonance", employing vibrato and glissando to create layered nuances in melody—examples include rolling vibrato, pressing vibrato, and multiple types of slide (Zhao Hanyang, 2003)[23]. The Erhu also features diverse special techniques, such as simulating natural sounds, as well as plucking and double-stopping to achieve different tonal qualities and expressive textures.

In contrast to Western bowed instruments, the Erhu exhibits fundamentally different characteristics in both construction and performance. It uses a snakeskin membrane as a resonator, and the bow is placed between its two strings; there is no fingerboard, so pitch is adjusted by varying finger pressure and sliding, resulting in a fluid, microtonal palette. In terms of musical aesthetics, the Erhu favors a single-line melodic approach with a strong focus on tonal nuance, conveying emotional tension through subtle timbral changes. These structural and aesthetic distinctions enable the Erhu to command a unique appeal on the global music stage.

2.2 Related Work

The methods for interaction design in NIME have evolved rapidly, offering musicians novel ways to express themselves through technology. These methods encompass various sensors, interfaces, and systems that bridge human gestures and sound production.

The most representative air instrument, Theremin[3], first appeared in 1920. Its playing style and technical implementation logic have inspired research related to hands free musical interfaces, including our works. Many people have also improved, explored, and developed new enhanced instruments based on the original framework of the Theremin, laying the foundation for gesture based interactive music.

Over the years, the development of new interaction technologies has opened up more possibilities for interactive musical performances. In 2021, Atau Tanaka systematically reviewed common techniques for capturing physiological information in gesture-based music interaction design[22].

In the realm of sensor technology, optical sensing—including motion capture (MoCap) systems—and devices such as Leap Motion enable users to interact with musical parameters and sound generation without requiring physical instruments or wearable devices, thus creating a more fluid and immersive musical experience. Within NIME, some have utilized Leap Motion for musical interface design[20] and augmented instrument design[7]. Likewise, IMUs, bending sensors, and pressure sensors have garnered significant attention in NIME research, primarily in wearable systems[3], augmented instruments[16], and DMI design[21]. Some researchers, such as Leimu[2], have even combined Leap Motion with IMUs to achieve more precise gesture tracking.

Over the past two years, as computer vision technology has advanced, methods that use algorithms such as Mediapipe[14] to capture spatial information have increasingly drawn attention. Although these vision-based approaches may not offer the highest precision, they eliminate the need for specialized hardware, making spatial gesture tracking more accessible. Ilya Borovik and colleagues (2023) developed a mobile web application that processes video streams to extract body motion data, enabling real-time interaction between movement and music[1].

Research into these various interaction methods—ranging from optical sensing to physiological interfaces—continues to expand the expressive potential of digital music performance. Each approach offers distinct advantages in expressivity, immersion, and ease of interaction, thereby pushing the boundaries of both live and studio music creation. As these technologies evolve, musicians will be able to engage with sound in increasingly innovative ways that closely reflect their physical movements.

2.3 Differentiable Digital Signal Processing

Differentiable Digital Signal Processing (DDSP) is a technique that integrates traditional DSP algorithms into a neural network training framework[5]. Originally proposed by Engel et al. in the context of a harmonic-plus-noise synthesizer for modeling monophonic and harmonic instruments, DDSP implements differentiable DSP modules within neural networks, enabling loss functions to perform backpropagation not only on model parameters but also directly on generated audio.

A prominent use case of DDSP is timbre transfer, wherein a model learns the timbral characteristics (such as time-varying harmonic amplitudes) of a particular instrument and then maps the frequency and amplitude information of other input sources—encompassing pitch and loudness control signals—to generate sounds that embody the target instrument' s timbral identity. Compared to traditional parameter-mapping or feature-engineering methods, DDSP better preserves the acoustic details of the target instrument and allows for flexible control over pitch or loudness during the inference stage. Jordie Shier et al. (2024) introduced a novel difference loss function for DDSP, enabling mappings from drum timbres to synthesizer parameters[17]. In the design of future DMIs and interactive music performance systems, DDSP is poised to offer even more possibilities.

2.4 Erhu in NIME

Within NIME and the broader field of human-computer interaction, research related to the Erhu remains relatively

[2] https://en.wikipedia.org/wiki/Huqin
[3] https://en.wikipedia.org/wiki/Theremin

sparse. Existing work has primarily focused on employing various methods—such as computer vision algorithms[13], magnetic position sensors[10], and score recommendation or grading algorithms[9] —to build Erhu learning environments. However, the development of digital instruments and interactive music performance centered on the Erhu is still a field with limited exploration to date.

3. TECHNICAL IMPLEMENTATION

The creation of the instrument involves three stages: dataset recording, model training, and interactive design. We first recorded an original dataset of Erhu performances, trained a DDSP model using this dataset, and then built specific performance rules and an interface in Max.

3.1 Dataset Collection

During the recording process, we performed Erhu improvisations in a stable and controlled recording environment, ensuring that the notes covered the common playing range of the Erhu, from the lowest D4 to D7, spanning three octaves. We aimed to showcase various long-tone techniques of the Erhu, including legato, glissando, vibrato, trills, and bowing techniques, at different dynamic levels. This approach ensured that the tonal characteristics of the Erhu were captured for each performance technique.

For the dataset, we used two microphones, an AT2020 and a TOPPING CL101, for A/B testing. After recording, we obtained two 20-minute mono WAV files (48kHz, 32-bit) as the raw dataset. This dataset will be made available to researchers and artists alike[4].

3.2 Model Training

The DDSP model training process consists of two stages: data preprocessing and model training.

3.2.1 Data Preprocessing

The DDSP library, developed by Magenta using TensorFlow[5] , is a differentiable processing framework and toolkit. During training, the raw WAV files must be converted to TFRecord format using the ddsp_prepare_tfrecord function from the DDSP library. Before conversion, the effects.trim function from the Librosa library[6] was used to remove portions of the audio with a volume below 20 dB, ensuring that the data processing would not overflow when applying a logarithmic transformation to the loudness.

After removing the silent sections, the audio files were processed using the following parameters:

Sample rate (–sample_rate=16000)

Frame rate (–frame_rate=50)

The audio was split into 4-second samples (example_secs=4.0) with a 1-second sliding window (hop_secs=1.0) to increase dataset diversity

If necessary, F0 (fundamental frequency) features were smoothed using the Viterbi algorithm (–viterbi=True)

Waveforms were center-aligned (–center=True)

The final TFRecord files were split into multiple shards (e.g., train.tfrecord-00000-of-00010), all stored in a designated data_dir folder.

3.2.2 Model Training

For model training, we set five training steps at 30,000, 35,000, 40,000, 45,000, and 50,000 iterations, resulting in five VST model files. Based on testing, we found that the model trained for 30,000 steps performed the best in terms of tonal fidelity and stability. The exported model files were placed in the preset folder of the DDSP-VST[7] plugin developed by Magenta. These models were then invoked using the External VST in the Max environment to perform subsequent tone conversion tasks.

3.3 Interactive Design

We developed interactive performance systems in Max.

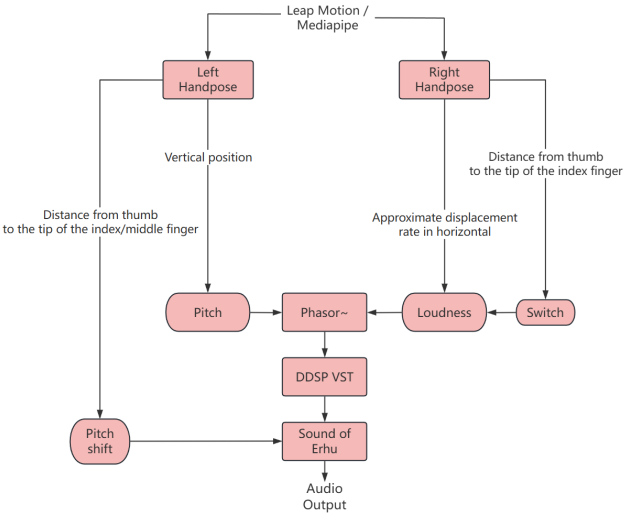

Figure 1: The interactive design process for the entire performance system.

3.3.1 Data Input

The performance system provides two options for capturing spatial information of the hand model: a)Using the Mediapipe algorithm to extract 2D hand model coordinates from camera images. b)Using Leap Motion to capture 3D spatial coordinates of the hand model. When using Mediapipe for hand model information, we developed a local Python script that interfaces with the current camera device and the Mediapipe library. The extracted hand model data is transmitted to the performance system via OSC (Open Sound Control)[8] protocol for real-time interaction.

For input via Leap Motion, we utilized an external object *Ultraleap*[9] developed by Jean-Michaël Celerier to facilitate communication between Leap Motion and Max. While the spatial data from Mediapipe is less accurate and has higher latency compared to Leap Motion, it does not require additional hardware, offering a lower-cost option for users to interact with the performance system.

3.3.2 Data Mapping

The interactive design logic is inspired by real Erhu performance techniques. The goal is to make the instrument performance as humanistic as possible while minimizing difficulty and learning costs. To achieve this, we used the spatial position data of both hands as input to control the pitch

[4] https://github.com/MINNE-WU/Gesture-Driven-Erhu
[5] https://www.tensorflow.org/
[6] https://librosa.org/
[7] https://magenta.tensorflow.org/ddsp-vst
[8] https://ccrma.stanford.edu/groups/osc/index.html
[9] https://github.com/celtera/ultraleap

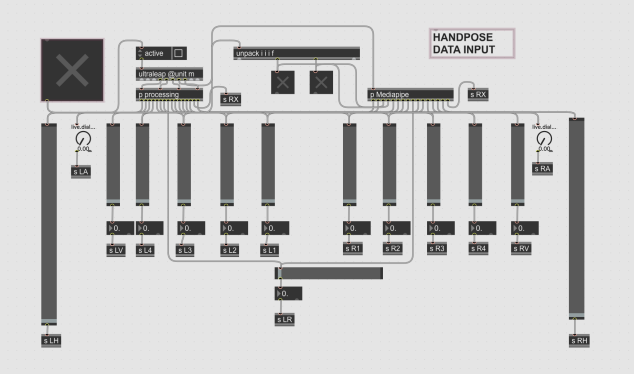

Figure 2: The original hand model data within the Max project is processed and then input into the interactive performance system. This data includes the spatial vertical and horizontal positions of both hands, the speed of movement along the vertical spatial direction, the distance between the thumb and other fingertips, the distance between the two thumbs, and the rotation angle of both hands around the forearm axis. This data not only drives the performance of the Erhu timbre but also fully utilizes these factors to build a complete interactive performance system.

and loudness of a triangle wave generated by a Phasor object, which is then sent to the DDSP VST for real-time conversion into the authentic sound of Erhu playing.

We mapped the vertical position of the left hand in space to control the pitch of the instrument. The system offers two modes: Free Pitch Mode and Quantized Pitch Mode. In Free Pitch Mode, the pitch changes freely without being constrained to a specific scale. In Quantized Pitch Mode, the pitch snaps to the nearest note in the selected scale, creating discrete tonal steps that make performance easier. In Quantized Pitch Mode, the user can adjust the ramp time of glissando from one note to another in milliseconds, enabling sliding between discrete notes. The adjustable ramp time range is from 0 to 500 ms. When ramp time is not 0, the player can still perform the glissandos of erhu playing.

Additionally, we computed the spatial distance between the left-hand index/middle fingers' tips and the thumb's tip. As this distance approaches zero (i.e., when the index/middle fingers are pinching the thumb), the pitch of the Erhu sound from the DDSP VST output is adjusted. When the index and thumb pinch, the pitch shifts up by a semitone; when the middle finger and thumb pinch, the pitch shifts up by a whole tone. This mimics the vibrato technique in Erhu playing and enriches the functionality of the performance system.

Inspired by the motion of the right hand holding the bow during Erhu performance, we calculated the horizontal spatial position of the right hand to control the instrument's volume. The value controlling the volume represents the difference in the position of the right hand along the horizontal axis between the current moment and 500 milliseconds earlier. This time interval is set deliberately to ensure the hand movement data is smooth and stable, as the hand model can sometimes show unstable or sudden jumps. By using this 500-millisecond interval, the displacement measured reflects the movement over that time window, allowing for a smoother and more stable response in volume changes. This method essentially captures the rate of change in the hand's position with a slight delay, incorporating a smoothing effect to reduce instability in the data.

To make the system's volume more controllable, we also calculated the spatial distance between the thumb and index fingertip of the right hand. The Phasor will only generate sound when the thumb and index fingertip are pinched together. When the thumb and index finger are separated, the volume of the Phasor will be set to zero.

We also built a User Interface (UI) for the performance system, with the final interface shown in the figure 3.

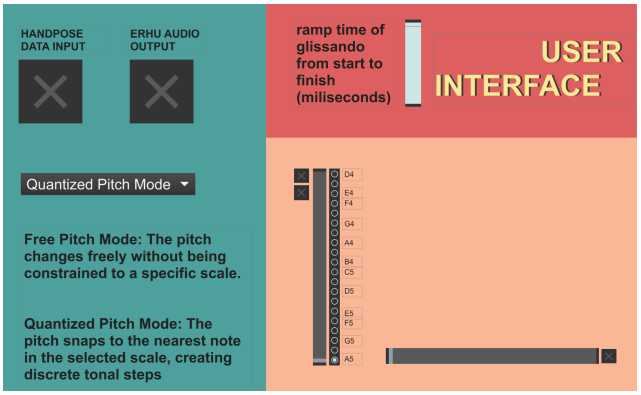

Figure 3: The User Interface of the performance system. The performer uses the vertical slider to determine the current pitch of the performance and the horizontal slider to adjust the volume of the performance system. The Performer can choose between two different pitch modes and set the ramp time for glissando under the quantized pitch mode.

4. PERFORMANCE

We tested this performance system in three different scenarios. The first performance was a duet, where we created a gesture-controlled Techno-style music system. One Person controlled the frequency band ratio and sound parameter changes with their hands, while the other person improvised over a stable rhythmic base and looping chords.

In the second performance, we recorded loops of our performance within 30-second cycles in DAW, layering them to create harmonic textures.

The third performance is the video we uploaded. We used an embodied interactive system, using the distance between various fingertips and the thumb to trigger random pitch and ADSR changes for pluck and pad synth sounds in phaseplant[10]. The speed of movement along the vertical axes of both hands triggered percussion sounds and neurobass (A Synthesizer Bass sound used in electronic music), and the rotation angle of the left hand controlled the wet/dry ratio of a particle effect. In addition to this, we performed the Erhu sound, aiming to create a rich and interactive performance.

As both the system's designers and experienced Erhu players, we felt authentic feedback while operating this performance system. The most noticeable aspect was that when the volume input to the DDSP changed, the output Erhu sound did not simply reflect a change in loudness; it mimicked the effect of varying bowing pressure in real Erhu playing, with the tone of the instrument shifting according to the intensity of the bowing.

It's worth mentioning that that the CREPE pitch tracking model used in the original DDSP is based on well-

[10]`https://kilohearts.com/products/phase_plant`

tempered chromatic scales of Western Classical Music. Therefore, there are potential difficulties of timbre-mapping instruments outside of this tradition. However, in our system of playing music according to the equal temperament, the tone of the erhu can still be well represented in the scale.

Overall, the system's tonal expression was humanistic and controllable, and the handling of spatial parameters for the gesture model allowed for further possibilities in performance and interaction design.

5. DISCUSSION

5.1 Limitations

While the gesture-controlled Erhu system presented in this study demonstrates promising results in terms of interaction and sound synthesis, several limitations must be acknowledged. First, the accuracy and precision of the spatial gesture tracking, especially when using the Mediapipe algorithm for hand recognition, can be affected by environmental factors such as lighting and background interference. The Leap Motion system offers more precise tracking but requires additional hardware, which may not be accessible to all users. Consequently, the system's usability can be constrained by these limitations in gesture recognition, particularly in non-ideal conditions.

Another limitation lies in the DDSP model itself. While the DDSP-based Erhu model preserves the unique timbre of the traditional instrument, there remain challenges in perfectly capturing the nuanced dynamics of Erhu performance, such as variations in bowing pressure and subtle finger positioning. Despite careful data collection, certain expressive qualities, such as the spontaneous variations in vibrato or pitch bend, can be difficult to emulate fully with the current model. Further exploration of more advanced modeling techniques, including better feature extraction and more refined training datasets, could help address these gaps.

5.2 Future Work

In terms of future work, one possible direction is to enhance the gesture recognition system, exploring alternative technologies such as advanced motion capture systems or multi-sensor fusion approaches that combine optical sensing with inertial measurement units (IMUs). This could improve accuracy while retaining the accessibility benefits of the current system.

Furthermore, the current DDSP model could be extended to better capture the dynamics of different Erhu performance techniques. For example, future models could incorporate temporal and dynamic modeling of various bowing techniques and the fine nuances of left-hand finger movements. Research into using more sophisticated loss functions and neural network architectures could help to better model these complex relationships between gestures and sound.

Another avenue for future research could involve expanding the interactive design to include multi-instrumental integration, allowing for real-time control of other traditional Chinese instruments within the same system. This could create a more immersive experience where different instruments interact in real-time, thus enhancing the overall digital performance.

6. CONCLUSIONS

This study has presented a novel approach to integrating traditional Erhu performance techniques with modern digital technologies, specifically using DDSP and gesture recognition. By mapping the spatial information of hand gestures to sound synthesis parameters in real-time, the proposed system provides a new way of performing and interacting with the Erhu in a digital context. Through the use of machine learning and differentiable signal processing, we have successfully preserved the unique timbral qualities of the Erhu while enabling intuitive control over its sound.

Although there are challenges to overcome, particularly in the accuracy of gesture recognition and the complexity of modeling the full expressive range of the Erhu, this work offers a significant step toward bridging traditional Chinese instruments with modern technological practices. The contributions of this study not only enhance the understanding of how traditional music can be integrated into the digital realm but also offer valuable insights for future developments in digital musical instrument design, providing a foundation for further research into the intersection of culture, technology, and music. Through continued exploration, we hope to contribute to the global evolution of traditional Chinese instruments, offering new possibilities for their digital adaptation and cultural expression.

7. ACKNOWLEDGMENTS

We would like to express our sincere thanks to Prof. Raul Masu, who guided us in writing the Mediapipe script, designing the embodied interactions, and writing the paper; Zichong XU, who helped us to record the Erhu dataset; and Chenxi BAO, who provided technical support. Thanks to NIME reviewers for their time and feedback.

8. ETHICAL STANDARDS

The datasets used in the model training process were all played by the authors. All related photos and videos were taken by the authors, and no unauthorized assets were used in this project.

References

[1] Ilya Borovik and Vladimir Viro. 2023. Real-time co-creation of expressive music performances using speech and gestures. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

[2] Dom Brown, Nathan Renney, Adam Stark, Chris Nash, and Tom Mitchell. 2016. Leimu: Gloveless music interaction using a wrist mounted leap motion. *Interaction* 1, 2 (2016), 3–4.

[3] Doga Cavdir, Romain Michon, and Ge Wang. 2018. The BodyHarp: Designing the intersection between the instrument and the body. In *Proceedings of the 15th International Conference on Sound and Music Computing*. Limassol, Cyprus.

[4] S. Emmerson. 2018. *The Routledge research companion to electronic music: Reaching out with technology*. Taylor Francis.

[5] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable digital signal processing. *arXiv preprint arXiv:2001.04643* (2020).

[6] S.H. Guo. 2019. *Laizi Zhongguo de shengyin: Zhongguo chuantong yinyue gailan [Sounding China: A companion to Chinese traditional music]*. Shanghai Yinyue Chubanshe, Shanghai.

[7] Jihyun Han and N. E. Gold. 2014. Lessons learned in exploring the Leap Motion sensor for gesture-based instrument design. In *Proceedings from Goldsmiths University of London*.

[8] Y. Hui and J. P. J. Stock. 2023. *The Oxford handbook of music in China and the Chinese diaspora*. Oxford University Press.

[9] Gwo-Haur Hwang, Ping-Tsung Tsai, Jenn-Kaie Lain, and Shiuan-Han Huang. 2023. Development and usability evaluation of an intelligent personalized Erhu pitch and rhythm learning system. In *Proceedings of the International Conference on Computers in Education*.

[10] Fumitaka Kikukawa, Sojiro Ishihara, Masato Soga, and Hirokazu Taki. 2013. Development of a learning environment for playing Erhu by diagnosis and advice regarding finger position on strings. In *NIME*, Vol. 2013. 271–276.

[11] C.F. Liu. 2018. *Erhu xitong jinjie lianxiqu ji [Erhu systematic progressive exercises]*. Zhongyang Yinyue Xueyuan Chubanshe, Beijing.

[12] Y.H. Liu. 1997. *Liutianhua quanji (Diyiban) [The complete works of Liu Tianhua (First edition)]*. Renmin Yinyue Chubanshe, Beijing.

[13] Bonnie Lu, Chyi-Ren Dow, and Chang-Jan Peng. 2020. Bowing detection for Erhu learners using YOLO deep learning techniques. In *HCI International 2020-Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*. 193–198.

[14] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[15] Y.M. Qian. 2020. *Piaomiao guhongying - Abing yanjiu wenji [The shadow of the ethereal and lonely shadow–A Bing's research anthology]*. Suzhou Daxue Chubanshe, Suzhou.

[16] Courtney N Reed, Adan L Benito, Franco Caspe, and Andrew P McPherson. 2024. Shifting ambiguity, collapsing indeterminacy: Designing with data as Baradian apparatus. *ACM Transactions on Computer-Human Interaction* 31, 6 (2024), 1–41.

[17] Jordie Shier, Charalampos Saitis, Andrew Robertson, and Andrew McPherson. 2024. Real-time timbre remapping with differentiable DSP. *arXiv preprint arXiv:2407.04547* (2024).

[18] Atau Tanaka, David Fierro, Francesco Di Maggio, Martin Klang, and Stephen Whitmarsh. 2024. The eavi EMG/EEG board: Hybrid physiological sensing. *arXiv preprint arXiv:2409.20026* (2024).

[19] Atau Tanaka, Federico Visi, Balandino Di Donato, Martin Klang, and Michael Zbyszyński. 2023. An end-to-end musical instrument system that translates electromyogram biosignals to synthesized sound. *Computer Music Journal* 47, 1 (2023), 64–84.

[20] Daniel Tormoen, Florian Thalmann, and Guerino Mazzola. 2014. The composing hand: Musical creation with Leap Motion and the BigBang Rubette. In *NIME*. 207–212.

[21] Sam Trolland, Alon Ilsar, Ciaran Frame, Jon McCormack, and Elliott Wilson. 2022. AirSticks 2.0: Instrument design for expressive gestural interaction. In *NIME 2022*. PubPub.

[22] Federico Ghelli Visi and Atau Tanaka. 2021. *Interactive machine learning of musical gesture*. Springer International Publishing, Cham, 771–798. `https://doi.org/10.1007/978-3-030-72116-9_27`

[23] H.Y. Zhao. 2003. *Erhu jifa yu mingqu yanzou tishi [Erhu technique and tips for playing famous pieces]*. Renmin Yinyue Chubanshe, Beijing.