

MODELING PERCEPTUAL LOUDNESS OF PIANO TONE: THEORY AND APPLICATIONS

Yang Qu^{1,2,*} Yutian Qin^{1,3,*}

Lecheng Chao¹

Hangkai Qian¹

Ziyu Wang^{1,4}

Gus Xia^{1,4}

¹ Music X Lab, NYU Shanghai

² City University of Hong Kong

³ New York University

⁴ Mohamed Bin Zayed University of Artificial Intelligence

yangqu7-c@my.cityu.edu.hk, {yq2120, lc4087, hq443, ziyu.wang, gxia}@nyu.edu

ABSTRACT

The relationship between perceptual loudness and physical attributes of sound is an important subject in both computer music and psychoacoustics. Early studies of “equal-loudness contour” can trace back to the 1920s and the measured loudness with respect to intensity and frequency has been revised many times since then. However, most studies merely focus on synthesized sound, and the induced theories on natural tones with complex timbre have rarely been justified. To this end, we investigate both theory and applications of natural-tone loudness perception in this paper via modeling piano tone. The theory part contains: 1) an accurate measurement of piano-tone equal-loudness contour of pitches, and 2) a machine-learning model capable of inferring loudness purely based on spectral features trained on human subject measurements. As for the application, we apply our theory to *piano control transfer*, in which we adjust the MIDI velocities on two different player pianos (in different acoustic environments) to achieve the same perceptual effect. Experiments show that both our theoretical loudness modeling and the corresponding performance control transfer algorithm significantly outperform their baselines.¹

1. INTRODUCTION

Sound *intensity* and *loudness* are two relevant but very different terms. Intensity is the physical feature of sound derived from sound pressure. Loudness, however, is the perceptual measure dependent on our auditory systems, where other physical factors also contribute to human perception. For example, a 1 kHz tone is measured louder than a 100 Hz tone of equal intensity, but softer than an equal-intensity white noise. Previous psychoacoustic studies provide knowledge of loudness perception via human subject experiments and various computational models have been proposed to account for loudness estimation [1–6].

However, existing theories mainly focus on synthesized

tones and thus fall short of explaining natural tone loudness such as the piano tone. The generation of piano tone involves a complicated physical process [7] and the sound contains rich timbral variations hard to be synthesized from both frequency and time domain perspectives [8, 9]. On the other hand, recently we see a growing number of music information retrieval tasks involving feature extraction of piano tone loudness, such as automatic music transcription [10–12] and performance rendering [13, 14]. In most of these studies, loudness is sometimes confused with intensity or even the MIDI velocity. This motivates us to investigate loudness perception specific to piano tone, beneficial for various downstream applications.

In this paper, we investigate both theory and applications on loudness perception specific to piano tone. The theory is studied in a hybrid method containing a *psychoacoustic* procedure and a *machine-learning* procedure. In the psychoacoustic procedure, we measure the “equal-loudness contour” on piano based on a human subject experiment similar to the pure-tone equal-loudness measurement [1]. Since piano tones cannot be directly controlled by frequency and intensity, we instead study pitch and velocity, two corresponding discrete performance controls. In our experiment, the controls are executed by (acoustic) player pianos for accuracy and reproducibility. We show that different pianos (in different acoustic environments) have distinctive equal-loudness contour patterns, and existing methods have limited power to explain the measured pattern.

In the machine-learning procedure, we extend the piano-tone equal-loudness contour to a computational model, capable of inferring loudness purely from spectral features. Traditionally, loudness models are based on mathematical approximation of our auditory systems [3–6]. We argue such an approach is laborious and even intractable for natural tones such as the piano tone to take into account all contributing facets of sound. Instead, we propose a machine-learning approach including a loudness model trained on the human subject measurement in the previous experiment. The model takes in the spectrogram and outputs the estimated loudness, calibrated to standard loudness unit *sones* in post-processing.

^{*}The two authors have equal contribution.

¹Code and models can be accessed via <https://github.com/yangqu2000/ModelingPerceptualLoudnessOfPianoTone>



We show our theory of piano tone loudness provides a more reasonable explanation of piano sounds in downstream applications. Specifically, we apply our loudness model to *performance control transfer* [14], in which we adjust the MIDI velocities on two different player pianos (in different acoustic environments) to achieve the same perceptual effect. The original intensity-based loudness estimator is replaced with the proposed method, and experimental results show a significant improvement in terms of transfer quality.

2. RELATED WORK

The study of the equal-loudness contour (ELC) of pure tone is a serious subject in psychoacoustics, since it reveals fundamental facts of human loudness perception with respect to frequency spectrum and intensity. The experiment settings have been modified throughout the century, including the tuning of listening conditions [1, 15, 16], the improvement of experiment procedures [1, 17, 18], and the extension of measured frequency range [1, 16, 19, 20]. Suzuki *et al.* [2] provides a detailed literature review of the existing experiments. Our piano-tone ELC experiment is modified from the original one, including careful adjustment to account for discrete frequency and intensity control.

The psychoacoustic listening tests of pure tone and many others [1, 16–18, 21, 22] provide fundamental understanding of our auditory systems. In return, the study of *loudness models* takes in the auditory system hypotheses and yields loudness estimation of the different types of tone, including [1, 16–18] for pure tone, [21] for complex tone, and [22, 23] for complex tone with amplitude modulations. So far, the state-of-the-art loudness model ISO 532-3 [24, 25] has the theoretical power to estimate the loudness of arbitrary waveforms. In our paper, we validate and compare with this model particularly on piano tone loudness estimation.

The majority of research on piano tone mainly focuses on the relationship between instrument control and the physical attributes of the generated sound, either statistically [26–28] or via physical modeling [7, 29, 30]. However, there is a lack of formal theory to cover the perception of piano tone. In this paper, we provide the first approach to study the relationship between instrument control and piano tone loudness in a psychoacoustic approach.

3. PIANO TONE EQUAL-LOUDNESS CONTOUR

Unlike pure tone which is determined by frequency and intensity, piano tone is largely dependent on the environment including the instrument itself that generates the sound (e.g., upright or grand piano) and the surrounding acoustic environment (e.g., concert hall or small room). Given a fixed environment, a piano tone can be controlled by four factors, namely *pitch*, *velocity*, *duration* and *pedal* [31]. Simple as they are, the four control parameters interact with the acoustic environment and produce a wide spectrum of piano timbre, resulting in an unexplored loudness perception pattern. In this paper, we study how perceptual loudness is affected by the first two factors, i.e., pitch

and velocity, which can be roughly understood as the “frequency” and “intensity” of piano tone, respectively. We leave the others for future study.

The piano tones in our experiment are produced by player pianos, which can execute the control parameter in a more accurate manner than human pianists. Pitch is controlled by 88 MIDI pitches in [21..108] and velocity is controlled by 128 velocity levels in [0..127].

3.1 Method of Measurement

The goal of this experiment is to measure the *equal-loudness contours* (ELC) of piano tone on a specific piano and acoustic environment, that is, a sequence of piano keys under possibly different velocities that have equal perceptual loudness.

The experiment is modified from the original pure-tone ELC experiments [1]. Specifically, we first select a reference tone, denoted by $(p_{\text{ref}}, v_{\text{ref}})$, where p_{ref} is a reference pitch and v_{ref} is the velocity level that we are interested in. Then, we enumerate piano pitches and for each pitch p_{var} , we ask subjects to listen to $(p_{\text{var}}, v_{\text{var}})$, $v_{\text{var}} \in [0..127]$ and choose the louder note until they find a velocity v_{var}^* such that $(p_{\text{var}}, v_{\text{var}}^*)$ and $(p_{\text{ref}}, v_{\text{ref}})$ have equal loudness.

In psychoacoustic terms, the reference tone $(p_{\text{ref}}, v_{\text{ref}})$ is called a *reference stimulus*, and given the pitch, the candidate tones under multiple velocities to compare (i.e., $(p_{\text{var}}, v_{\text{var}})$) are called *variable stimuli*. The solution $(p_{\text{var}}, v_{\text{var}}^*)$ is called the *point of subjective equality* (PSE). The curve that connects the PSEs at multiple pitches is the equal-loudness contour at reference velocity v_0 .

To adjust the velocity of reference stimuli and search for the PSE of each pitch, we adopt *randomized maximum likelihood sequential procedure* (RMLSP) [1], a common procedure used in pure-tone ELC experiments consisting of a series of test rounds. At each test round, the subject is asked to compare the loudness of a pair of piano tones, including the reference stimulus and a variable stimulus. Then we fit an online logistic regression model to predict the PSE according to the subject’s response so far. The velocity of the variable stimulus in the next test round is randomly selected within a velocity range centered at this PSE velocity. The algorithm ensures the procedure will converge to the correct PSE.

3.2 Experiment Setting

In our experiment, we use A4 in the middle of the keyboard as the reference pitch (i.e., $p_{\text{ref}} = 69$). We measure equal loudness contours at four velocity levels: $v_{\text{ref}} \in \{32, 44, 60, 80\}$. The four velocities lie in the common range of velocity usage and are evenly distributed with respect to the statistical distribution [11].

We measure equal-loudness contours on 9 variable pitches: $p_{\text{var}} \in \{21, 33, 45, 57, 69, 81, 93, 105, 108\}$. The pitches are the A’s in all octaves together with the highest note C8. The measurement on all 88 keys is ideal though not affordable, since RMLSP normally takes 20 minutes to find the PSE with respect to one single variable pitch.

ID	Environment		Number of subjects assigned to each velocity level				
	Instrument	Acoustic environment	$v_{\text{ref}} = 32$	$v_{\text{ref}} = 44$	$v_{\text{ref}} = 60$	$v_{\text{ref}} = 80$	Total
Env. I	Grand Disklavier	Anechoic chamber	6	6	6	5	23
Env. II	Upright Disklavier	Non-anechoic chamber	7	7	7	7	28

Table 1: Experiment settings and subjects assignment in two environments.

We address the loudness modeling on the other pitches in section 4.

We set the number of RMLSP test round to be 32. In the first two test rounds, the subject always listens to two piano tones of constant variable velocities $v_{\text{var}} = 90$ and $v_{\text{var}} = 30$, respectively. After that, the model yields the next variable velocity uniformly sampled from a $[-6..+6]$ interval centered at the current-step estimation of the PSE. The order of the piano tone pairs between the reference stimulus and the variable stimulus is random at each test round.

The experiment is conducted on two player pianos located in different acoustic environments (as shown in Table 1). Environment I contains a grand YAMAHA Disklavier piano in an anechoic chamber, and Environment II contains an upright YAMAHA Disklavier piano in a non-anechoic chamber. Subjects are seated half a meter in front of the pianos where the pianists usually sit and use a laptop to respond.

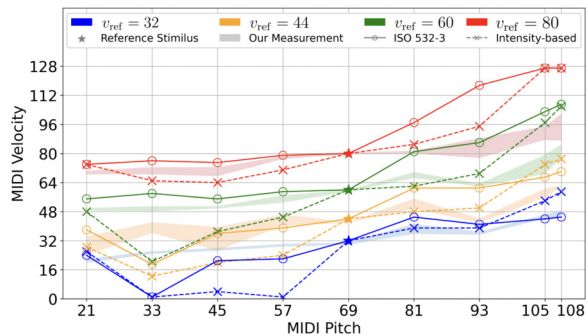
The subjects are 15 adults (18–35 years), 10 male and 5 female. All subjects have normal hearing sensitivity. 11 subjects participate in the experiment in Environment I and 9 subjects participate in the experiment in Environment II. Each subject is tested in a separate section.

The subject is tested at two or three of the four velocity levels randomly assigned (as shown in Table 1). For Environment I, 23 measurement results were collected. Each velocity level has a sample number of 6 except for velocity 80 with a sample number of 5. For Environment II, 28 measurement results were collected. Each velocity level has a sample number of 7.

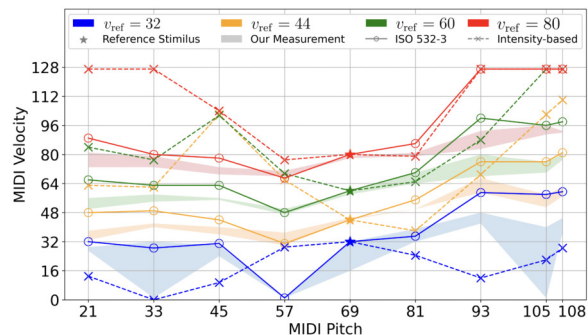
3.3 Result

Figure 1 shows our experiment results in Environment I and Environment II, respectively. Since MIDI velocity is an ordinal variable, it is improper to average the subjects' PSEs and show the mean equal-loudness curves. Instead, we present *equal-loudness ribbons*, where we replace the mean with the range of subjects' PSEs between the first and third quartiles among all data. Here, we use four different colors to indicate four reference velocities, and the diamond markers indicate the four reference stimuli on each graph. Note that in the lowest ($v_{\text{ref}} = 32$) ribbons, some of the variable pitches have PSEs equal to one, meaning that the variable pitches at the least audible velocity are still louder than or equal to the reference tone.

We see the patterns of equal-loudness ribbons are very different across the two environments. In Environment I, we see a gradual growth trend in all velocity levels, meaning higher pitches of the same velocity tend to be softer.



(a) Environment I.



(b) Environment II.

Figure 1: Measurement of equal-loudness ribbons of piano tone in two environments compared with ISO 532-3 [24,25] and intensity-based loudness computation [14,26].

In Environment II, the trend is less evident, and we see a valley at A3 (i.e., $p_{\text{var}} = 57$) at all reference velocity levels. Moreover, the range of equal-loudness ribbons varies, indicating the just noticeable difference in velocity change is uneven on different pianos, pitches and velocities.

3.4 Comparison with Existing Methods

We use the experiment result to validate two common loudness computation methods. The first method is ISO 532-3 by Moore *et al.* [24,25], the state-of-the-art loudness model based on the modeling of human auditory systems. The loudness value is computed as the maximum value of the predicted long-term loudness curve. The other method is used in [14,26] by naively computing the average intensity of the first 10 ms after the peak. For the two methods, we use the recordings of all the piano tones (discussed in section 4.3) and compute their corresponding loudness. The induced equal-loudness contours are shown in Figure 1 in solid lines with circles and dashed lines with crosses, respectively. Other popular methods such as applying A-

weighting is not considered since such methods are proved not generalizable to natural tone [32].

As shown in Figure 1(a) and Figure 1(b), in both environments, the results induced from ISO 532-3 are in general consistent with the equal-loudness measurement except for the pitch in the higher registers. The results induced from the intensity-based method fluctuate around the equal-loudness ribbons, failing to describe the perceptual effect.

4. LOUDNESS MODEL

In the previous section, we measure the equal-loudness contours (ELC) in two environments on 9 variable pitches and 4 reference velocities. In this section, we extend our findings to learn loudness models to account for loudness estimation of the remaining tones via a non-parametric approach as baseline (discussed in section 4.1) and a parametric approach (discussed in section 4.2). The parametric model is also capable of predicting loudness given the waveform of an arbitrary piano tone. We compare the prediction result with ISO 532-3 and the intensity-based method in section 4.3.

4.1 Non-parametric Method

We first propose a naive approach to predict the loudness of arbitrary piano tones in each environment via linear interpolation of the measured ELCs in section 3. First, we assign the loudness in *sones* to the tones of the reference pitch A4 under all the velocities using the calibration method discussed in section 4.2.3. Then, we linearly interpolate the ELCs measured on 9 variable pitches to all 88 pitches, and assign the same loudness level along each contour. Finally, for each pitch, we linearly interpolate between the ELCs. In section 5, we see such a simple method already yields satisfactory performance in downstream applications.

4.2 Parametric Method

Moreover, a parametric model is proposed to estimate the loudness using machine learning.

4.2.1 Model

We use $x = (p, v)$ to denote a piano tone, where p is the pitch and v is the velocity levels. We learn a parametric loudness model $\ell = f_{\theta}(x)$ to compute the loudness of a given piano tone x .

The loudness model is learned as a supervised classification problem (as shown in Figure 2). The input is a pair of piano tones (x_1, x_2) , where $x_1 = (p_1, v_1)$ and $x_2 = (p_2, v_2)$. The target is the ground-truth label y , which is the indicator of whether x_1 is the louder one inferred from the ELCs. The difference between the estimated loudness ℓ_1 and ℓ_2 is used to predict the y . Specifically, the loss function is:

$$\mathcal{L}(\theta; x_1, x_2, y) = \text{BCE}(f_{\theta}(x_1) - f_{\theta}(x_2), y), \quad (1)$$

where $\text{BCE}(\cdot)$ is the binary cross entropy function.

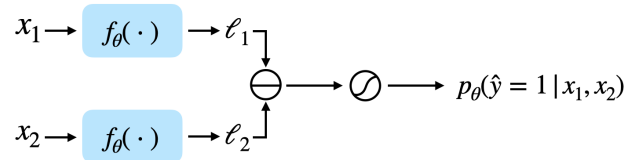


Figure 2: Illustration of the objective function.

The pairs (x_1, x_2) are sampled from all the tone pairs in each environment whose loudness comparisons are derivable based on the human subject results. Specifically, a pair must satisfy either of the two conditions:

- (C.1) The two tones have the same pitch and thus can be compared based on velocity. The one with the larger velocity is assumed to be louder.
- (C.2) The two tones have different variable pitches and the loudness comparison can be inferred from the equal-loudness ribbons in Figure 1.

4.2.2 Implementation

We define our loudness model $\ell = f_{\theta}(x)$ as the linear combination of the values on the mel-spectrogram of x (without an intercept term). The mel-spectrogram has 8 mel-frequency bins, a 2048 window size, and a 512 hop size under the sample rate of 22050 Hz. We select the 5 time frames (approx. 0.1s) right after the onset of the tone detected by [33]. In all, we have 80 input features of a pair of piano tones and the model can be optimized using logistic regression.

The small mel-frequency bin number in the current model can be understood as a constant convolution kernel to achieve better generalization on other non-variable pitches. In the future, the model can also be extended to neural networks given a larger amount of human subject results.

4.2.3 Calibration

The parametric method learns an unstandardized estimation of piano tone loudness based on classification, which needs to be further calibrated to the standard loudness unit in *sones*. We address the problem by matching the piano tones under reference pitch A4 to the corresponding equal-loudness pure tones and compute the corresponding pure tone loudness using ISO 532-3. The matching procedure follows from the *method of adjustment* [32].

4.3 Evaluation

4.3.1 Data Preparation

We record the piano tones of two Disklavier pianos in both environments for all pitches and velocities. The recorder is placed 0.5 meters in front of the piano (similar to the subjects' ear position) and is kept still during the recording procedure. For all piano tones, the MIDI duration is 0.3 seconds and each recording clip lasts 1.3 seconds.

To train the parametric model, we select all the piano tone pairs that satisfy the two conditions (defined in 4.2.1),

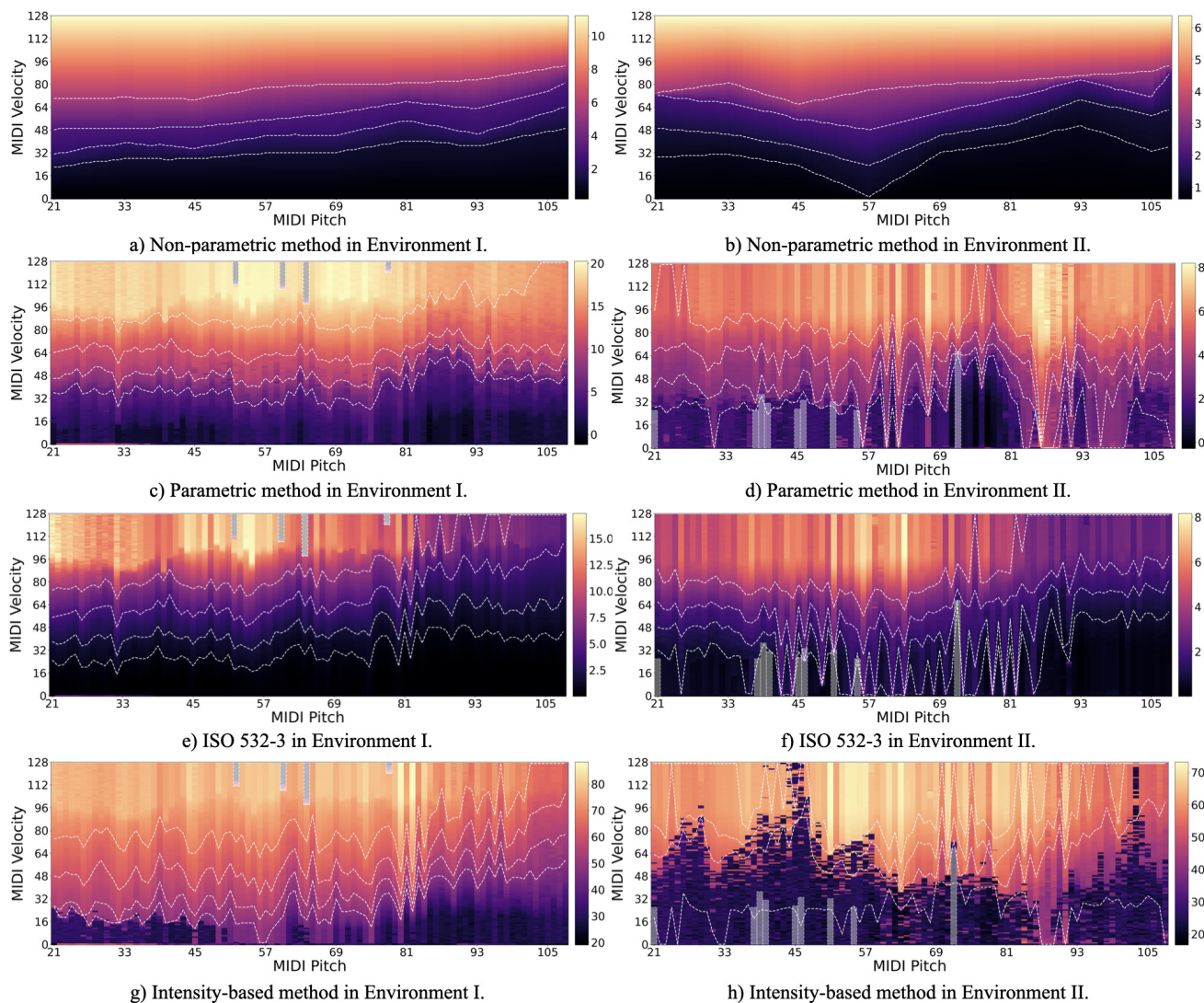


Figure 3: Visualization of calculated loudness value in Environment I and Environment II using our proposed loudness models, ISO 532-3 [24,25], and intensity-based method [14,26]. Points with a rectangular mask indicate mechanical failure of the player piano.

including 673,035 pairs satisfying the **C.1** and 427,633 pairs satisfying the **C.2** in Environment I, and 695,860 pairs satisfying the **C.1** and 422,386 pairs satisfying the **C.2** in Environment II. The dataset is randomly split into train set (80%) and test set (20%).

4.3.2 Results

Figure 3 shows the heatmaps of the predicted loudness of all piano tones in both environments by four methods: 1) our non-parametric method, 2) our parametric method, 3) ISO 532-3, and 4) intensity-based method. We show the ELCs derived from the heatmaps in dashed lines for better readability.

We see the non-parametric method presents a smooth interpolation of the measured ELCs. The result of the parametric method demonstrates a similar trend as ISO 532-3 except in the higher pitch range, which is more consistent with human subject measurement in Figure 1. Moreover, both parametric method and ISO 532-3 show more discon-

tinuity than Figure 1, suggesting the MIDI velocity is not assigned in a uniform manner for different pitches. Finally, the result of intensity-based method are noisy and inconsistent with the human perception.

Table 2 shows the model accuracy on the test set of two conditions defined in 4.2.1 separately. The results demonstrate that our proposed method outperforms the two baselines in both environments, especially in the accuracy of **C.2**. We show the generalization ability in two ways. First, we apply the model trained in one environment to the other environment and achieves satisfactory accuracy (as indicated by the underlined numbers). Second, we combine the training data in both environments and train a hybrid model. The model outperforms the baselines and even achieves the best performance in environment II in both conditions. The evaluation result shows our proposed model capture certain features in the piano-tone mel-spectrogram contributing to the perceptual loudness.

Methods	Acc. in Env. I		Acc. in Env. II	
	C.1	C.2	C.1	C.2
ISO 532-3	0.9689	0.9631	0.9037	0.9455
Intensity-based	0.9658	0.9290	0.8377	0.8555
PM - Env. I	0.9689	0.9893	<u>0.8976</u>	0.9614
PM - Env. II	<u>0.9528</u>	<u>0.9607</u>	0.9121	0.9793
Hybrid PM	0.9401	0.9785	0.9370	0.9881

Table 2: Evaluation of loudness comparison accuracy in Environment I and Environment II. The methods to compare are: 1) ISO 532-3: [24,25], 2) Intensity-based [14,26], 3) PM - Env. I: Parametric model trained on the data recorded in Environment I, 4) PM - Env. II: Parametric model trained on the data recorded in Environment II, and 5) Hybrid PM: Parametric model trained on the data recorded in both environments. Underlined data are tested by the parametric model training in the other environment.

5. PERFORMANCE CONTROL TRANSFER

In this section, we apply our theory of piano tone loudness to the downstream application of *performance control transfer*, in which we adjust the MIDI velocities on two different player pianos to achieve the same perceptual effect. An example scenario of the application is to reproduce a pianist’s performance in the concert hall to one’s living room.

5.1 Modification

Performance control transfer is originally proposed in [14], in which piano tone loudness is treated as a fundamental invariant property between the original performance and the transferred version. However, the study mistakenly uses the physical intensity as the loudness measure. Our approach replace the loudness estimator by the proposed loudness models (discussed in section 4) and keep the remaining algorithm the same. We also use ISO 532-3 as the loudness estimator as a baseline method.

5.2 Listening Test

We conduct a listening test to invite participants to evaluate the performance transferred by different algorithms, similar to the one conducted in [14]. Specifically, we prepare four pieces, including two monophonic pieces and two polyphonic pieces. The pieces are selected to cover classical and popular genres.

We invite two pianists to play the four pieces in Environment I and record both pianists’ MIDI control and the performance audio. Then, we transfer the MIDI files recorded in Environment I to Environment II using different performance transfer methods, including 1) our non-parametric method, 2) our parametric method, 3) ISO 532-3, 4) intensity-based model (i.e., the original transfer method [14]), and 5) a raw transfer without any MIDI velocity editing. We play the transferred MIDI in Environment II and record them from the same microphone position discussed in section 4.3.

We invite people to subjectively rate the transfer quality through a double-blind online survey. During the survey,

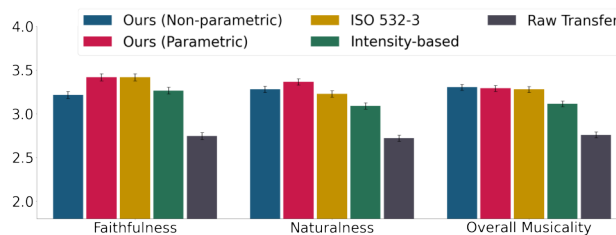


Figure 4: The subjective evaluation results of the five transfer methods.

the subjects listen to four groups of samples. In each group, the original performance in Environment I is played, followed by the five transferred versions. Both the order of groups and the sample order within each group are randomized. After listening to each sample, the subjects rate them based on a 5-point scale from 1 (very low) to 5 (very high) according to three criteria: *faithfulness* (to the original performance), *naturalness* and *overall musicality*.

5.3 Results

A total of 21 participants (14 male, 7 female) with different musical backgrounds have completed the survey. Figure 4 shows the result where the heights of the bars represent the means of the ratings and the error bars represent the confidence intervals computed via within-subject ANOVA [34].

The results show that all the loudness-based methods significantly outperforms the original intensity-based implementation and raw transfer (with p-value < 0.005). Moreover, our parametric method is marginally better than ISO 532-3 in terms of naturalness and overall musicality, and comparable with ISO 532-3 in faithfulness. Both our parametric and non-parametric methods demonstrate with only sparse human subject measurement data, we can achieve promising performance transfer results.

6. CONCLUSION

In this paper, we have contributed the first study to model the perceptual loudness of piano tones in a hybrid approach combining psychoacoustic experiments, and data-driven methods. Our theory include: 1) measurements of piano-tone equal-loudness contours on two player pianos, and 2) data-driven loudness models to estimate piano-tone loudness based on the measured contours. Experiments show our model provides more satisfactory estimation than existing loudness theories. Based on our findings, we validate the state-of-the-art loudness model ISO 532-3 on piano-specific tasks and argue that existing methods can be greatly improved if we have a more detailed human subject measurement and a deeper understanding of the timbral features of piano tone.

On the application side, we improve the existing performance control transfer method with the proposed loudness estimator. We believe the other downstream applications can also benefit from our study to propose a clearer problem definition and more accurate feature extraction.

7. REFERENCES

- [1] H. Takeshima, Y. Suzuki, H. Fujii, M. Kumagai, K. Ashihara, T. Fujimori, and T. Sone, "Equal-loudness contours measured by the randomized maximum likelihood sequential procedure," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 389–399, 2001.
- [2] Y. Suzuki and H. Takeshima, "Equal-loudness-level contours for pure tones," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 918–933, 2004. [Online]. Available: <https://doi.org/10.1121/1.1763601>
- [3] B. C. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of the Audio Engineering Society*, vol. 45, no. 4, pp. 224–240, 1997.
- [4] E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical band width in loudness summation," *The Journal of the Acoustical Society of America*, vol. 29, no. 5, pp. 548–557, 1957. [Online]. Available: <https://doi.org/10.1121/1.1908963>
- [5] B. R. Glasberg and B. C. Moore, "A model of loudness applicable to time-varying sounds," *Journal of the Audio Engineering Society*, vol. 50, no. 5, pp. 331–342, 2002.
- [6] H. Fletcher, "Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure," *The Journal of the Acoustical Society of America*, vol. 6, no. 2, pp. 59–69, 1934.
- [7] D. E. Hall and A. Askenfelt, "Piano string excitation v: Spectra for real hammers and strings," *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1627–1638, 1988.
- [8] B. Bank, F. Avanzini, G. Borin, G. De Poli, F. Fontana, and D. Rocchesso, "Physically informed signal processing methods for piano sound synthesis: a research overview," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 10, pp. 1–12, 2003.
- [9] B. Bank and L. Sujbert, "Generation of longitudinal vibrations in piano strings: From physics to sound synthesis," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2268–2278, 2005.
- [10] Q. Kong, B. Li, X. Song, Y. Wan, and Y. Wang, "High-resolution piano transcription with pedals by regressing onset and offset times," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3707–3717, 2021.
- [11] D. Jeong, T. Kwon, and J. Nam, "A timbre-based approach to estimate key velocity from polyphonic piano recordings." in *ISMIR*, 2018, pp. 120–127.
- [12] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: challenges and future directions," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 407–434, 2013.
- [13] A. Maezawa, K. Yamamoto, and T. Fujishima, "Rendering music performance with interpretation variations using conditional variational RNN." in *ISMIR*, 2019, pp. 855–861.
- [14] M. Xu, Z. Wang, and G. G. Xia, "Transferring piano performance control across environments," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 221–225.
- [15] B. Kingsbury, "A direct comparison of the loudness of pure tones," *Physical Review*, vol. 29, no. 4, p. 588, 1927.
- [16] H. Takeshima, Y. Suzuki, K. Ashihara, and T. Fujimori, "Equal-loudness contours between 1 khz and 12.5 khz for 60 and 80 phons," *Acoustical Science and Technology*, vol. 23, no. 2, pp. 106–109, 2002.
- [17] K. Betke, "New measurements of equal-loudness level contours," in *Proc. Inter-noise*, vol. 89, 1989, pp. 793–796.
- [18] J. Hall, "Maximum-likelihood sequential procedure for estimation of psychometric functions," *The Journal of the Acoustical Society of America*, vol. 44, no. 1, pp. 370–370, 1968.
- [19] H. Møller and J. Andresen, "Loudness of pure tones at low and infrasonic frequencies," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 3, no. 2, pp. 78–87, 1984.
- [20] H. Fasti, A. Jaroszewski, E. Schorer, and E. Zwicker, "Equal loudness contours between 100 and 1000 hz for 30, 50, and 70 phon," *Acta Acustica united with Acustica*, vol. 70, no. 3, pp. 197–201, 1990.
- [21] L. L. Beranek, J. Marshall, A. Cudworth, and A. P. G. Peterson, "Calculation and measurement of the loudness of sounds," *The Journal of the Acoustical Society of America*, vol. 23, no. 3, pp. 261–269, 1951.
- [22] B. Roß, C. Borgmann, R. Draganova, L. E. Roberts, and C. Pantev, "A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones," *The Journal of the Acoustical Society of America*, vol. 108, no. 2, pp. 679–691, 2000.
- [23] S. Kuwada, R. Batra, and V. L. Maher, "Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones," *Hearing research*, vol. 21, no. 2, pp. 179–192, 1986.

- [24] “Acoustics — Methods for calculating loudness — Part 3: Moore-Glasberg-Schlittenlacher method,” International Organization for Standardization, Geneva, CH, Standard, 2017.
- [25] B. C. J. Moore and B. R. Glasberg, “Modeling binaural loudness,” *The Journal of the Acoustical Society of America*, vol. 121, no. 3, pp. 1604–1612, 2007. [Online]. Available: <https://doi.org/10.1121/1.2431331>
- [26] R. B. Dannenberg, “The interpretation of midi velocity,” in *ICMC*, 2006.
- [27] A. Adli, Z. Nakao, and Y. Nagata, “Calculating the expected sound intensity level of solo piano sound in midi file,” vol. 2006, 2006, pp. 731–736.
- [28] M. Keane, “Statistical analysis of classical piano recordings in midi format,” in *Proceedings New Zealand Acoustical Society Conference*, 2004.
- [29] B. Bank and J. Chabassier, “Model-based digital pianos: from physics to sound synthesis,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 103–114, 2018.
- [30] S. Ewert and M. Müller, “Estimating note intensities in music recordings,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 385–388.
- [31] O. Ortmann, *The Physical basis of piano touch and tone: An experimental investigation of the effect of the player’s touch upon the tone of the piano.* K. Paul, Trench, Trubner & Company Limited, 1925.
- [32] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models.* Springer Science & Business Media, 2013, vol. 22.
- [33] B. McFee, A. Metsai, M. McVicar, S. Balke, C. Thomé, C. Raffel, F. Zalkow, A. Malek, Dana, K. Lee, O. Nieto, D. Ellis, J. Mason, E. Battenberg, S. Seyfarth, R. Yamamoto, viktorandreevichmorozov, K. Choi, J. Moore, R. Bittner, S. Hidaka, Z. Wei, nullmightybofo, A. Weiss, D. Hereñú, F.-R. Stöter, P. Friesch, M. Vollrath, T. Kim, and Thassilo, “librosa/librosa: 0.9.1,” Feb. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.6097378>
- [34] H. Scheffe, *The analysis of variance.* John Wiley & Sons, 1999, vol. 72.