SOUND PRODUCTION OF MIDI PIANO TONES IN A DB SCALE ON THE BASIS OF EQUAL LOUDNESS

TOMOYASU TAGUTI
Konan University
Department of Information Science and Systems Engineering
taguti@konan-u.ac.jp

Abstract : A method to control the velocity of piano tones of MIDI tone synthesizers on the basis of equal loundness property is presented. The idea is a parameterization of equal loudness contour of velocity with a physical correlate to the loudness. A listening experiment was conducted to get the equal loudness contour on a particular sound synthesizer with the method of paired comparison, where the standard stimulus was note C_4 and the comparison stimuli were notes n in the chromatic scale $C_2 \sim C_7$. Regarding the physical correlate to the loudness, the A-weighted sound pressure level (L_A) was chosen to fix the method. By relating the L_A-value, ℓ, of C_4 to the equal loudness contour, a mapping formula from (n, ℓ) to velocity v, i.e., $g : (n, \ell) \mapsto v = g(n, \ell)$, was established. By this mapping, note n played at velocity $g(n, \ell)$ produces a tone whose loudness is equal to the tone of C_4 sounded at ℓ dB in the sense of L_A. The formula was implemented as part of a computer-aided piano performance system to specify dynamic expression.

Key Words : *Midi velocity, Piano tone, Equal loudness contour, A-weighted SPL*

1 Introduction

The sound intensity (or sound volume) of MIDI-controlled musical instruments (MIDI instruments, for short, which in this paper include tone synthesizers without keyboard or the like) is specified with a 7-bit integer code called “velocity”. But the MIDI specification book gives the definition of its acoustical characteristic only in that the larger the velocity, the larger the volume of sound, with a suggestion to use a logarithmic relationship between the two[1].

This vague functional definition of velocity creates a problem in dynamic expression reproduced on different MIDI instruments, that is, the same MIDI message results in different responses in dynamics across different MIDI instruments.

Further, It seems rather common in the present-day MIDI instruments that the same velocity given to different notes, even among chromatically adjacent notes, does produce a different sound intensity (or volume, loudness, or what may be termed for similar auditory impression.) Manually adjusting such an uneven response to velocity leads to almost an endless task when synthesizing music performance with software.

This paper presents an experimental method to control the velocity of piano tones of MIDI instruments. The method is based on (1) the contour of equal loudness obtained by listening experiment, and (2) a physical correlate to the loudness to parameterize this contour.

Apparently the method cannot achieve a perfect control over the velocity in terms of desired loudness since the degree of resolution is limited due to the MIDI velocity system that allows only 127 steps. Further, the musical expression of dynamics is not identical to loudness of raw instrumental sound. Nevertheless, the method is expected to improve controllability of dynamics for synthesis of music performance with software.

The MIDI instrument dealt with in this paper is MU2000 from YAMAHA Corp. It is a MIDI tone synthesizer, one of general-purpose synthesizers without keyboard, that provides several timbre sets of piano tones and of other instrumental tones. The target timbre dealt with in this paper is "stretched tuning grand piano stereo" which seems to be suited for rendition of classical piano music; this is available with plug-in board PLG150-PF added to MU2000. The instrumental parameters were set to the initial, factory setting that turns off all the supplemental sound effects and equalizing functions. This instrument will be referred to as Module M hereafter.

2 Contour of equal loudness

2.1 Method of loudness matching

The standard stimulus was note C_4, and the comparison stimuli were the notes in the chromatic scale of $C_2 \sim C_7$. Let N denote the set of all the notes taken for stimuli, i.e.,

$$N = \{C_2, C_2\#, \cdots\cdots, C_7\} \tag{1}$$

The goal of the matching task is to determine the velocity of each of comparison stimuli so that its loudness is equal to that of the standard stimulus (C_4) produced with a given velocity. For ease of the experimental task, the nearest C from below was chosen as the substitute for C_4 when the comparison stimuli are higher by more than one octave from C_4. Similarly, the nearest C from above was chosen as the substitute for C_4 when the comparison stimuli are lower by more than one octave from C_4. The velocity of this substitute was set at the value that produced the loudness equal to C_4's. Thus, the notes in $C_5\# \sim C_6$ were matched to C_5 where C_5's loudness was equal to C_4's, and the notes in $C_6\# \sim C_7$ were matched to C_6 where C_6's loudness was equal to C_5's as determined by the above procedure. The notes in $C_2 \sim B_2$ were matched to C_3 where C_3's loudness is equal to C_4's.

(1) Subject: A female student majoring piano playing in a graduate school of music education.

(2) Listening environment: The experiment was conducted in a low reverberant listening room (volume 61 m^3, reverberation time 0.3 s at 500 Hz, background noise level 26 dBA). See Fig. 1 for the setup of the experiment. The output level of the stereo loudspeakers was set so that the sound of C_4 at velocity 127 becomes 90 dB SPL (fast) at the listening point of the subject. The subject was listening to the stimulus pair by adjusting the velocity by herself according to the procedure which follows.

(3) Stimuli: Let **A** and **B**, respectively, denote the standard and comparison stimuli produced with velocities V_A and V_B. Here, V_A took m integers from the set W = {5, 10, 20, 30, 40, 50, 60, 70, 80 ,90, 100, 110, 120, 127}, $m = 14$. For later use, let V_i ($i = 1, 2, \cdots, m$) denote these elements of W so that

$$V_1 = 5, V_2 = 10, \cdots, \text{ and } V_m(= V_{14}) = 127 \tag{2}$$

hence $W = \{V_i\}$, and $V_A \in W$. Velocity V_B took any integer from 1 through 127.

Stimuli **A** and **B** took a 500 ms duration each, being followed by a 500 ms silence. The duration 500 ms was chosen by taking into consideration the effect of temporal summation of loudness [2]. The silent interval 500 ms was chosen by following [3].

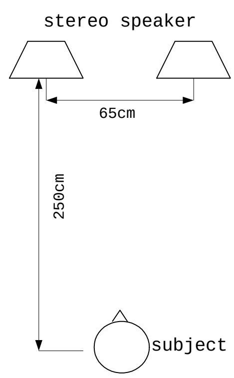

Figure 1: *Setup of the listening experiment.*

(4) Procedure: Six sequences of stimuli {**A A A A**}, {**B B B B**}, {**A B A B**}, {**B A B A**}, {**A B A B A B**}, and {**B A B A B A**} were provided, and the subject was able to listen to any of these as many times as she became confident. She was asked to determine the value of V_B with which the loudness of **B** becomes subjectively equal to that of **A** with V_A. Write the determined V_B as $\bar{V}_B$. The stimulus pairs were taken in the order (1), (2), (3), (4) and (5) as specified in Table 1 below.

Table 1: Order of matching task

	A	**B**
(1)	C_4 with V_A	C_4#, $\cdots$, C_5
(2)	C_5 with $\bar{V}_B{}^{\dagger 1}$	C_5#, $\cdots$, C_6
(3)	C_6 with $\bar{V}_B{}^{\dagger 2}$	C_6#, $\cdots$, C_7
(4)	C_4 with V_A	B_3, $\cdots$, C_3
(5)	C_3 with $\bar{V}_B{}^{\dagger 3}$	B_2, $\cdots$, C_2

†1) V_B determined by (1)
†2) V_B determined by (2)
†3) V_B determined by (4).

W was partitioned into W_1 = {20, 40, 60, 80, 100} and W_2 = {5, 10, 30, 50, 70, 90, 110, 120, 127}, and the experiment was conducted in two sessions, i.e., Task 1 with $V_A \in W_1$ for notes $C_3 \sim C_7$, which had 240 stimulus pairs, and Task 2 with $V_A \in W_2$ for notes $C_3 \sim C_7$, as well as with $V_A \in W$ for notes $C_2 \sim B_2$, which had 600 stimulus pairs in total. Task 1 aimed at figuring out an overall profile of equal loudness property of Module M with as much accurate data as possible by following the method of limits. The role of Task 2 was to complete the result of Task 1 over the entire gamut of $C_2 \sim C_7$ with $V_A \in W$; Task 2 was conducted in a less stringent matching procedure so that the subject was asked to report just one $\bar{V}_B$, instead of a complete range of $\bar{V}_B$'s for each stimulus pair.

2.2 Result

Among 240 pairs of Task 1, 209 pairs yielded a single integral number of $\bar{V}_B$, and 31 pairs yielded two consecutive integral numbers of $\bar{V}_B$. For the latter case, the average of the two consecutive integral numbers was taken as the effective $\bar{V}_B$ for further proceeisng.

The result in the portion of chromatic scale C_4# $\sim$ C_6 in Task 1 is shown in Table 2. It displays a general tendency that the higher the note, the larger the value of $\bar{V}_B$. However, it is not monotonic or smoothly changing, e.g., $\bar{V}_B$ of C_4# is smaller than V_A when V_A = 20, 40, 80 and 100, and $\bar{V}_B$ of D_4 is considerably larger than V_A when V_A = 40, 60, 80 and 100; this fact implies that C_4# is louder and D_4 is considerably softer than C_4 when these three notes are played as single tones with the same velocity 40, 80, or 100.

Table 2: Result of Task 1.

Shown are $\bar{V}_B$ of comparison stimuli in chromatic scale C_4# $\sim$ C_5 against $V_A \in W_1$.

C_4	20	40	60	80	100
C_4#	19	38	60	76	98
D_4	20.5	43	65.5	87.5	114
D_4#	22	44	64	88	111
E_4	22	41.5	62	84.5	105
F_4	21.5	41.5	66	86	109
F_4#	23	43	64	83	107.5
G_4	21.5	45	64	84	106
G_4#	23	47	66	86	108
A_4	23	47.5	63	85.5	105.5
A_4#	23	48	65.5	90.5	107
B_4	24	48	65	92.5	110
C_5	26	49	68	93.5	112
C_5#	23	44	62.5	84	107
D_5	27	47	65	90	111
D_5#	23	48	63	88	106
E_5	23	44	64	86	106
F_5	26	50	67	92	114
F_5#	24	49	68	94	112
G_5	28	48	69	95	114
G_5#	28	50	68.5	94	114
A_5	26	49	69	92.5	116
A_5#	31	51	69	90	113
B_5	30.5	55	72	95	112
C_6	29	55	72	96.5	113

Among 600 pairs of Task 2, 97 pairs could not yield a correct matching because of insufficient loudness of **B** with $V_B \leq 127$. A nominal value $\bar{V}_B = 127$ was given to these pairs.

At this point, let us change the notation V_A to V_i. (See (2).) The entire result of $v = \bar{V}_B$ (Tasks 1 and 2 together) is regarded as a function of two discrete variables n and V_i such that

$$v = f(n, V_i) \ \text{ for } n \in N \text{ and } V_i \in W \qquad (3)$$

Here, $f(n, V_i)$ is assumed to satisfy the identity mapping for $n = \mathrm{C}_4$ that

$$V_i = f(\mathrm{C}_4, V_i) \ \text{ for } V_i \in W \qquad (4)$$

Figure 2 shows the profile of $v = f(n, V_i)$ in a form of equal loudness contour, being parameterized with $V_i \in W$ ($i \leq 13$), in the note–velocity space.

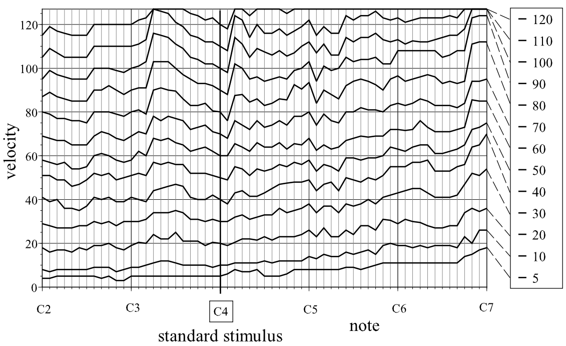

Figure 2: *Equal loudness contour of velocity $f(n, V)$ for $V = V_i$ ($1 \leq i \leq 13$), in the range $C_2 \sim C_7$.*

Introduce a function $\tilde{f}(n, V)$, V being a real variable V, that interpolates $f(n, V_i)$ with respect to the second variable. For simplicity reasons, we shall choose a piecewise linear interpolation, hence the formula is given as:

For $^\forall V \in [V_i, V_{i+1}]$ ($i = 1, 2, \cdots, m-1$),

$$\tilde{f}(n, V) = a \times f(n, V_i) + (1-a) \times f(n, V_{i+1}) \qquad (5)$$
$$\text{where} \quad a = (V_{i+1} - V)/(V_{i+1} - V_i).$$

The quantity $v = \tilde{f}(n, V)$ gives the velocity of note n (if v is not an integer, it must be rounded to the nearest integer) so that the loudness of note n produced with velocity v is (nearly) equal to that of C_4 with velocity V (integer) in $V_1 \leq V \leq V_m$. Hence, $\tilde{f}(n, V)$ is called the parameterization of equal loudness contour with the velocity of the standard stimulus (C_4).

3 Parameterizing the equal loudness contour in the scale of L_A

3.1 A-weighted SPL as an interval scale

Among the known physical correlates to the loudness of unsteady sounds, we chose the A-weighted sound pressure level (L_A) in *fast* response mode for the interval scale of loudness. In the following, however, the L_A-values are not the absolute ones but represent offsets measured from an arbitrary origin, and 'dB' always denotes the dB scale in this sense.

Let $p(t; n, V)$ be the electrical output signal of single piano tone of note n at velocity V. Actually, the analog, stereo tones at a fixed output level were sampled digitally at 48 kHz sampling, and the recorded waveforms were mixed down to monoral singals to yield $p(t; n, V)$. Let $p_A(t; n, V)$ be the A-weighted waveform of $p(t; n, V)$.

Put

$$L_{n,V} = \max_{t_0 \leq t \leq t_1} 10 \log_{10} J_{n,V}(t) \qquad (6)$$

where

$$J_{n,V}(t) = \int_{-\infty}^{t} |p_A(s; n, V)|^2 e^{-(t-s)/\tau} \, ds/\tau \qquad (7)$$

Here, the time constant $\tau = 125$ ms is due to the *fast* response mode, and t_0 and t_1, respectively, are time points before and enough after the onset of the target tone. Figure 3 displays the entire profile of L_{n,V_i} for $V_i \in W$ over all 88 notes in a form of equal velocity contour.

This graph gives us a useful information for understanding the acoustical aspect of sound production of Module M. The following are a few remarks on Table 3 together with Fig. 3. The dynamic range of Module M is 55 dB or a little more over the entire registers. Also, the contour lines display an uneven response of L_A against notes, where the magnitude of uneveness depends on V_i almost monotonously. This fact suggests that Module M generates very likely the tones of different velocities by processing (mainly) the

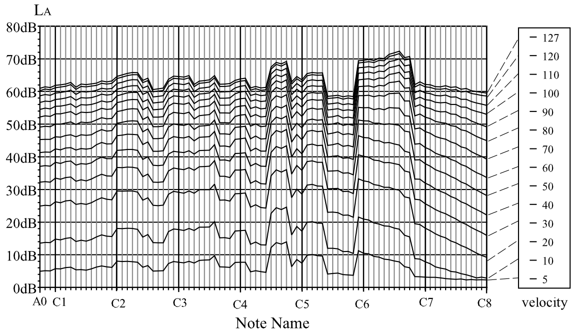

Figure 3: Equal velocity contour of L_A for $V = V_i$ ($1 \leq i \leq 14$) over 88 notes. (Note that the origin is set arbitrarily.)

amplitude of just a single reference waveform for each note.

Now, let us focus on the L_A-value of note C_4, L_{C_4,V_i}, writing as

$$L_i = L_{C_4,V_i} \tag{8}$$

Table 3 lists the values of L_i.

Table 3: Listing of $L_i = L_{C_4,V_i}$ dB for $V_i \in W$.

V_i	5	10	20	30	40	50	60
L_i	7.72	17.39	28.74	35.97	41.08	45.50	49.16

V_i	70	80	90	100	110	120	127
L_i	52.74	55.49	57.76	59.85	61.61	63.14	63.82

Introduce an interpolation function

$$\ell = \hat{h}(V) \tag{9}$$

on L_i so that it is monotone increasing in real variable V satisfying

$$L_i = \hat{h}(V_i) + L_{C_4,\hat{V}}, \quad V_i \in W \tag{10}$$

Here, $\hat{V}$ is some constant taken suitably as a referential velocity, e.g., $\hat{V} = 30$.

The function $\ell = \hat{h}(V)$ represents the L_A-value (of C_4) at velocity V relative to the one at velocity $\hat{V}$. We remark that, with a particular choice of $\hat{V} = 30$, the range of $\hat{h}(V)$ nearly balances around the origin, i.e., -28.25dB $\leq \hat{h}(V) \leq$ 27.85dB for $V_1 \leq V \leq V_m$, and $\hat{h}(\hat{V})$ =0 dB.

3.2 Formula of parameterization

Take the inverse mapping of (9)

$$\hat{h}^{-1} : \ell \mapsto V = \hat{h}^{-1}(\ell) \tag{11}$$

and put $\hat{h}^{-1}(\ell)$ into the second argument of $\tilde{f}(n, V)$:

$$v = g(n, \ell) \overset{def}{=} \tilde{f}(n, \hat{h}^{-1}(\ell)) \tag{12}$$

The function $v = g(n, \ell)$ gives the velocity v of note n whose loudness is (nearly) equal to that of C_4 at ℓ dB. (If v in not an integer, it must be rounded to the nearest integer.)

Numerical procedure to get v is given as follows when a piecewise linear interpolation is assumed for both $\tilde{f}(n, V)$ and $\hat{h}(V)$.

1. Get i that satisfies $\ell_i \leq \ell < \ell_{i+1}$ where $\ell_i = \hat{h}(V_i)$
2. Compute $a = (\ell_{i+1} - \ell)/(\ell_{i+1} - \ell_i)$
3. Compute $v = a \times f(n, V_i) + (1 - a) \times f(n, V_{i+1})$

Figure 4 shows the contour of $v = g(n, \ell)$ plotted for $\ell = 0, \pm 5, \cdots, \pm 25$ dB, in case of $\hat{V} = 30$.

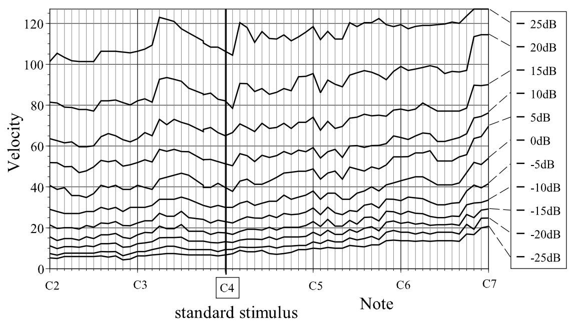

Figure 4: *Equal loudness contour of velocity in terms of* ℓ. *Shown are curves for* $\ell = 0, \pm 5, \cdots, \pm 25$ *dB when* $\hat{V} = 30$.

4 Conclusions and discussion

A method to control the velocity of MIDI piano tones with the L_A-value of a reference note (i.e., standard stimulus) as its interval scale was presented.

Because of limited degree of resolution in the MIDI velocity system, the method would not be able to provide a perfect result. Nevertheless, it gives us a practical, quantitative way of control over the velocity of MIDI piano tones with acoustically well-defined variable.

The given interpolation formula was implemented as part of a software of piano music performance on Module M. It enabled the user (the author) to realize his desired dynamic expression, e.g., smooth change of crescendo/diminuendo, good dynamic balance between melody and accompanying parts, vivid expression of trills etc., on a solid, acoustical basis [4], [5], [6]. This fact suggests that the proposed method may be used as a means for quantitative study on "performance rules" of dynamic expression in the domain of analysis in music performance.

From the psychoacoustic viewpoint, there remains a problem whether the L_A is a reasonable choice as the interval scale to parameterize the equal loudness contour. In the previous report [7], the author adopted the continuous A-weighted SPL, L_{Aeq}. The L_{Aeq} requires temporal integration, and a 500 ms window was chosen in order to make it equal to the duration of the stimuli for loudness matching; the fact that it is the duration of one beat played at tempo 120 bpm (beats per minute), a moderate tempo for typical melody lines to play, was another reason for this choice. The L_A (in *fast* response mode, maximum) and L_{Aeq} (with a 500 ms window of temporal integration) were found to have a similar dependence characteristics on the velocity for Module M, in the sense of $\ell = \hat{h}(V)$. However, a more detailed study seems necessary on the choice of interval scale.

As was mentioned, the measurement of acoustical variable as a function of velocity gives us a precise understanding on the response characteristic of sound production of MIDI instruments, and this understanding will lead us to a good starting point for preparing listening experiments on loudness matching. In this direction we are doing measurements on piano tones of other MIDI instruments, e.g. [8], which dealt with the profile of the L_A value and the spectral center of gravity of a MIDI-controlled acoustic piano.

References

[1] "MIDI-1.0 Standardization Book", *MIDI Kikaku Kyogikai*, 1980. (In Japanese)

[2] J. J. Zwisloski, "Temporal summation of loudness: An analysis", *J. Acoust. Soc. Am.*, **46**(2), 431–441 (1969).

[3] W. Reichardt and H. Nisse, "Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level", *J. Acout. Soc. Am.*, **47**(4), 1083–1090 (1970).

[4] T. Taguti, "Expression of motion in piano performance: A quantitative study with Mozart's piano sonata K. 311", *Proceedings of IAEA2002 (17th Congress of the International Association of Empirical Aesthetics, 4-8 August 2002, Takarazuka, Japan)*, 469–472 (2002).

[5] T. Taguti, "Rendering Chopin's Waltz No.6 and Mozart's Piano Sonata No. 17 (1st Movement) with system MUSE ; A technical note", *Workshop Program APP-5, IJCAI2003, 11 August 2003, Acapulco, Mexico*, 82–84 (2003).

[6] T. Taguti, "Creating expressiveness with software: The case of Mozart's Piano Sonata No.18, D major", *Proceedings of ICA2004 (18th International Congress on Acoustics, 5-9 April 2004, Kyoto, Japan)*, V-3407–3410 (2004).

[7] T. Taguti, "Mapping a physical correlate of loudness to the velocity space of MIDI-controlled piano tones", *Proceedings of SMAC03 (Symposium on Music Acoustics 2003, August 6–9, 2003, Stockholm, Sweden)*, 565–567 (2003).

[8] T. Sawada and T. Taguti, "Sound pressure level and the spectral variation of 88 piano tones on a Disclavier", *Proceedings of ISMA2004 (International Symposium on Musical Acoustics 2004, March 31 – April 3, 2004, Nara, Japan)*, 211–214 (2004).