INTERNATIONAL SYMPOSIUM on MUSICAL and ROOM ACOUSTICS

ISMRA 2016

September 11-13, 2016

La Plata, Buenos Aires, Argentina

Music Perception: Paper ISMRA2016-9

Assessing the acoustic similarity of different pianos using an instrument-in-noise test

Alejandro Osses Vecchi[a], **Antoine Chaigne** [b], **Armin Kohlrausch**[a,c]

[a]Human-Technology Interaction group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, the Netherlands, a.osses@tue.nl

[b]Institute of Music Acoustics, University of Music and Performing Arts, Vienna, Austria

[c]Brain, Behaviour & Cognition group, Philips Research Europe, Eindhoven, the Netherlands

Abstract:

Speech-in-noise tests are a way to evaluate the perception of one or more aspects of phonetic elements. When varying the amount of noise, i.e., the signal-to-noise ratio (SNR), the target condition at which correct answers are given 50% of the times is of special interest. The obtained SNR is called speech reception threshold (SRT). In order to quantify how close two (non-speech) sounds are in terms of their acoustic properties, we hypothesise that a similar procedure can be followed. If we let subjects distinguish two sounds presented together with a background noise, and measure the SNR at which the two can no longer be distinguished, we expect a strong correlation between SNR and similarity: high similarity needs only little noise (high SNR) to make the sounds indistinguishable, whereas in a decreasing similarity an increasing amount of noise is needed (lower SNR). For a given pair of sounds, we can establish thus similarity by measuring the SNR at which they become indistinguishable. In this paper we present an example of such a method, applied to recorded Viennese piano sounds. As in every speech-in-noise test, the noise to be used has to be carefully chosen. We chose a set of noises that are shaped according to the spectro-temporal properties of each note and instrument. Such a set was generated in a similar way as the so-called ICRA noises in speech, but being adapted to provide a more suitable "piano weighting". Results for similarity scores for one piano note (C#$_5$) and 7 test pianos are presented and discussed.

Keywords: Musical instruments, music perception, listening tests, ICRA noise, signal-to-noise ratio.

INTERNATIONAL SYMPOSIUM on MUSICAL and ROOM ACOUSTICS
ISMRA 2016
September 11-13, 2016
La Plata, Buenos Aires, Argentina

Assessing the acoustic similarity of different pianos using an instrument-in-noise test

1 Introduction

Nowadays the piano represents one of the most popular instruments and its versatility has made it popular in different music styles as, for instance, classical music, rock, jazz and fusion. The piano was invented in the 18^{th} century, undergoing major construction changes during the 19^{th} century and being produced in a large scale during the 20^{th} century [1]. Although the piano has been produced for more than 200 years a complete understanding of its acoustics has not yet been reached.

Since its creation, the piano has kept the following main constitutive parts: hammer, string, bridge and soundboard. In brief, sound is generated when a key of the piano is pressed and the hammer strikes a string. This results in a propagating pulse that is transmitted to the soundboard, with the bridge as connecting element. The soundboard vibrates, being responsible for the radiation of sound. Although the constitutive parts of the piano have not changed during the years, their design has actually changed. These changes have affected the intrinsic characteristics of the piano. Some of these changes are: hammer mass and velocity, string tension and diameter and the soundboard thickness and rigidity. In a recent study, Chaigne et al. [2] showed how the amplitude (envelope), spectral content and decay time of a given piano note ($C\#_5$) was affected by changing the parameters just mentioned. The structural changes they introduced to the piano notes were simulated in an advanced computational model that accounts for the interactions between the constitutive parts [3, 4]. They pointed out that pianos constructed during the first half of the 19^{th} century had a reduced hammer mass, reduced string tension and thinner soundboard, compared to newer pianos. These outcomes motivated us to perform a similarity study considering Viennese pianos from the first and second half of the 19^{th} century (built between 1805 and 1873), where this transition in the structural construction of the piano occurred. A comparison of three of these pianos but from a physical point of view is presented in a recent study by one of the authors [5].

The aim of our study was not only to evaluate which pianos are more or less similar among each other, but also to develop a method to quantify how far or close this acoustic difference is. During this study two experimental methods were used: (a) the method of triadic comparisons, and; (b) a discrimination task in background noise. The method of triadic comparisons provides a way to convert perceptual similarity scores obtained from a preference test into a space where the Euclidean distance is inversely related to similarity [6, 7, 8]. A preference ranking of piano sounds using this method is used to validate the second method: the discrimination of piano sounds in a background noise. We developed this method with the hypothesis of being a measurable way to evaluate the acoustic similarity of pairs of sound. The similarity is then measured as a function of the signal-to-noise ratio (SNR) between the test piano sounds and the amount of noise needed to recognise two different sounds 70.7% of the time. In this case, the SNR represents a unidimensional measure that is proportional to the acoustic similarity between a given pair of sounds.

We start this paper by describing the details of the noise used in the identification task, followed by an example of how to use a combination of those noises to compare a given pair of piano sounds. Additionally, a detailed description of the two experimental procedures is given. Then, the results of the experimental sessions with 14 subjects are presented and both methods are compared. Although our main goal is to use the results of the perceptual similarity experiments to assess the impact of the construction differences between pianos, only indications about how to use this information is given. This paper is rather focused on the description and validation of the instrument-in-noise identification task.

2 Description of the method

In order to compare different piano sounds, we looked for a method able to give us a measurable indication about how acoustically similar two sounds are. When two sounds are very different their discrimination should be easy and, likewise, the discrimination should be more difficult if two sounds are more similar. When presenting three sounds, with one of them being different (target sound), our hypothesis is that its identification becomes gradually more difficult when the task is done with a background noise. This noise, however, has to be carefully chosen in order to produce the desired effect. Since the piano sounds have strongly varying temporal properties in addition to a rich spectrum, the noise to be used has to follow similar spectro-temporal properties. In a different context, the International Collegium of Rehabilitative Audiology (ICRA) developed an algorithm to generate a random noise that follows the spectro-temporal properties of a given speech signal [9]. We modified this algorithm to fit it to the case of the piano. In the following section we present a description of the modified algorithm.

2.1 ICRA noise with piano weighting

The procedure to generate the ICRA noises introducing a “piano weighting” is shown in Figure 1 and can be summarised as follows:

1. **Band split filter**: An input signal (piano sound) is split into 31 bands with centre frequencies between 87 Hz (3 ERB) and 7819 Hz (33 ERB). The total RMS level L_{tot} of the signal is stored for the level adjustment of the resulting noise.
2. **Sign shift**: the sign of each sample of the 31 filtered signals is either reversed or kept unaltered with a probability of 50% (multiplication by 1 or -1) [10]. As a consequence of this randomisation, the output samples have a flat spectrum.
3. **Re-filtering per band-split filter**: the resulting signal from band i is fed into the i-th band of the Gammatone filter bank. The index i represents each of the 31 bands.
4. **Add signals together**: the 31 filtered signals are added together.
5. **Phase randomisation**: the phase of the signal is randomised following a normal distribution from 0 to 2π, this is done in the frequency domain by overlapping/adding the segments after an inverse FFT with a 87.5% overlap. The resulting signal is adjusted to have the same total RMS level as in the band split filter stage.
6. **Low-pass filter at 8200 Hz**: a Butterworth filter, 8^{th} order with a cut-off frequency at the upper limit of the highest critical band was added to reduce the “scratchy sound” ($f_{\text{cut-off}}$ at 8200 Hz$\approx$33.5 ERB).

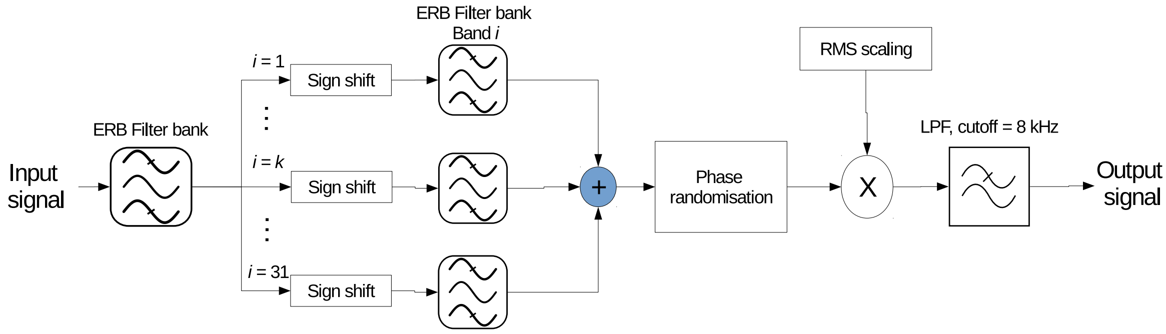


Figure 1: The principle of the ICRA noise generation. For details in the procedure, refer to steps 1 to 6 in the text.

2.2 Comparing two sounds using the ICRA noise

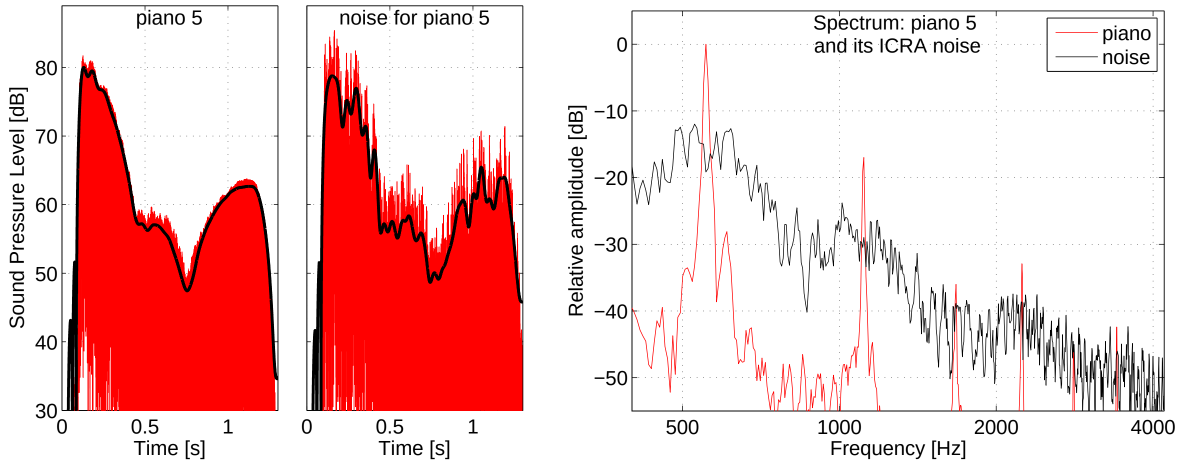
In this section we give an example about how to use the concept of ICRA noise to compare two piano sounds. We use two recordings of the note C#$_5$ (nominal frequency of 554 Hz) of the pianos 5 and 7 (see Table 1 for more details of the pianos). First, the ICRA noise for both pianos has to be generated individually. These ICRA noises have the same RMS level as the piano sounds and therefore are assumed to represent a signal-to-noise (SNR) ratio of 0 dB. As an example of this process, the waveforms of piano 5, its resulting ICRA noise and their spectra are shown in Figure 2. For the comparison, it is important that the duration of the piano samples and noises is the same, in this case set to 1.3 s, and that the piano onset is synchronised to occur at the same time stamp, in this case set to 0.1 s. Next, a joint ICRA noise (with an SNR$=0$ dB) has to be generated by combining the resulting ICRA noises (arithmetic mean between waveforms). In order to perform the actual comparison between piano 5 and 7 the SNR of the joint ICRA noise has to be adapted by applying a positive gain for reaching negative SNRs (noise is louder than the sounds) or a negative gain for reaching positive SNRs (sounds are louder than the noise). It is assumed that the resulting ICRA noise will be efficient to gradually mask the properties of both piano sounds (either piano 5 or 7) when presented together as the noise level increases (and the SNR decreases).

For each piano pair, twelve ICRA noises were generated. Within every trial of our 3-AFC experiment (see section 3.5), three of those noises were randomly chosen and presented together with the piano sounds. The use of the ICRA noises in this way corresponds to an approximation to a running noise condition.

3 Experimental method

3.1 Participants

Fourteen participants (8 females and 6 males) were recruited from the JF Schouten subject database of the Eindhoven University of Technology. The age of the participants ranged from 20 to 38 years old (average of 24 years old) and they all had self-reported normal hearing. They provided their informed consent before starting the experimental session and were paid for their contribution.



(a) Waveform of piano 5 (left) and its corresponding ICRA noise (right). The thick black lines correspond to the envelope of the waveforms.

(b) Spectra of piano 5 (red line) and its resulting ICRA noise (black line) averaged over the whole duration of the sounds.

Figure 2: Waveform of the Viennese piano 5 (a, left) converted to Sound Pressure Level and of its resulting ICRA noise at an SNR$=0$ dB (a, middle panel). In panel b the spectra of the piano sound and the ICRA noise are presented. See Table 1 for more details about piano 5.

3.2 General methods

The experiments were conducted in a doubled-walled sound-proof booth and they were organised in two one-hour sessions per participant. The piano sounds were presented through headphones in a diotic reproduction (left and right channels were the same). The participant's responses were collected on a computer after each trial presentation using the APE Toolbox for MATLAB [11] and the software APEX [12] for the triadic comparison and the instrument-in-noise experiments, respectively.

3.3 Stimuli

Recordings from seven pianos are compared among each other. Only one recording per piano was chosen leading to a total of 7 stimuli. The note C#$_5$ (nominal frequency $f_0 = 554$ Hz) was chosen. The total duration of the piano sounds was set to 1.3 s, with the note onset occurring at time 0.1 s. The sounds were ramped down using a 150-ms cosine ramp. The loudness of the piano sounds was set to a maximum value of 18 sone. For this purpose the short-term loudness obtained from the Time-Varying Loudness (TVL) model [13] was used. After loudness balancing the waveforms of the individual piano sounds had a maximum level ranging from 73 to 84 dB SPL.

By comparing 7 stimuli it is possible to construct 35 triads (7 choose 3) or 21 pairs of sounds (7 choose 2).

3.4 Procedure: triadic comparisons

The method of triadic comparisons consists of the presentation of three sounds “A”, “B” and “C”. After listening to all of them as many times as needed, the participant has to indicate

ID #	Short name	Manufacturer	Year of construction	Additional information
1	GH05	Gert Hecher	1805*	
2	CG28	Conrad Graf	1828	
3	JBS36	Johann Baptist Streicher	1836	
4	JBS51-4486	Johann Baptist Streicher	1851	Hammer action 4486
5	JBS51-4544	Johann Baptist Streicher	1851	Hammer action 4544
6	JBSS73	Johann Baptist Streicher & Sohn	1873	
7	NS19	Nannette Streicher	1819	

Table 1: List of pianos used in the perceptual comparison. The piano #1 has been built recently by Gert Hecher as a replica of a piano built in 1805.

the pair that is the most similar and the pair that is most dissimilar. With this information it is also possible to infer the pair with medium similarity. For instance, if the pair A-B is judged as the most similar and A-C as the most dissimilar, then a medium similarity is attributed to the remaining pair B-C. Each participant judged all 35 possible triads once. Within the 35 triads, each pair of piano sounds is judged 5 times.

3.5 Procedure: instrument-in-noise identification

The piano sounds were tested pairwise. For each pair of piano sounds the discriminability threshold was estimated by using an adaptive procedure (staircase method) with the level of the ICRA noise as adjustable parameter. Refer to section 2 for a description of the ICRA noises and how they were applied for each piano pair. A three alternative forced-choice (3AFC) paradigm was used (two sounds being the same, one being different), the participant had to identify which interval corresponded to the different piano sound. The variable parameter was adjusted following a two-down one-up rule (noise is increased after 2 consecutive right answers and decreased after 1 wrong answer). This paradigm tracks the 70.7% discriminability threshold. The trials for each pair of sounds were presented until 12 reversals were reached. The step sizes were set to 4 dB, 2 dB (after the second reversal) and 1 dB (after the fourth reversal). The threshold in dB was then assessed as the median value of the noise in the last 8 reversals. To avoid the use of loudness cues in the identification task, in addition to the loudness balancing of the stimuli, the presentation level of each interval (piano + noise) within the trials was randomly varied (roved) by ± 4 dB, drawn from a uniform distribution. We provided explicit instructions to not use level as identification criterion. Every participant provided 10 or 11 threshold estimates, meaning that by every two participants the 21 pairs were tested once.

4 Results

4.1 Triadic comparison

Considering the 7 test stimuli and the 35 possible triads, each of the 21 piano pairs is presented 5 times. A similarity matrix was constructed in order to summarise how often each of the 21 pairs was chosen as most similar, most dissimilar or indirectly chosen as having middle similarity. Each row i and column j in the matrix is a “similarity count” related to the piano pair

	Piano ID #						
Piano ID #	1	2	3	4	5	6	7
1	-	81	57	74	113	74	56
2		-	36	81	48	36	132
3			-	86	67	132	16
4				-	95	79	57
5					-	90	36
6						-	24
7							-

Table 2: Similarity matrix for participants S01-S15 (excluding S09). One presentation of all possible triads. Maximum score of 140 for each pair (2 x 5 pairs presentation x 14 participants).

i-j. If we consider a triad composed by pianos i, j and k with the pair i-j indicated as most similar and i-k as the most dissimilar, then the pair j-k is assumed to have middle similarity. This means that in the similarity count for this example: (a) will be increased by 2 units for row i, column j; (b) will be unchanged (0 units) for row i, column k, and; (c) will be increased by 1 unit for row j, column k. Since each pair is presented 5 times within all possible triads and 14 participants took part of the experiment, the maximum total score is 140 (2 times 5 times 14). The resulting similarity matrix obtained from the triadic comparisons is shown in Table 2. A similar procedure to construct a similarity matrix has been used in [6, 7, 8].

The similarity matrix was used as input for the classical multidimensional scaling (MDS) algorithm available in the MATLAB Statistics toolbox. MDS is commonly used as a visualisation tool of complex data. Each entry of our $n \times n$ similarity matrix is assigned to a lower-dimensional space ($n \times q$ matrix), where the distance between "treatments" ($n =$7 pianos) is related to the perceptual similarity among them. To find this reduced space, the eigenvectors ($n \times n$) and eigenvalues λ_i ($n \times 1$ matrix) corresponding to the dissimilarities scores $D_{i,j}$ are calculated, taking the q eigenvectors corresponding to the largest q eigenvalues. One criterion to test the adequacy of a $q-$dimensional representation is given by Equation 1, where a value P_q of at least 80% is considered to produce an adequate fit of the data in the q-dimensional space.

$$P_q = 100 \cdot \frac{\sum_{i=1}^{q} |\lambda_i|}{\sum_{i=1}^{n} |\lambda_i|} \tag{1}$$

For interested readers, a more detailed review of this theory is provided in [14].

In our case, after normalising the similarity matrix $S_{i,j}$ to the maximum score (140 units), the dissimilarity matrix D was then obtained by computing $D_{i,j} = \sqrt{1 - S_{i,j}}$ for each element i, j.

The obtained perceptual space has 4 dimensions with a goodness-of-fit of 99.5% (Equation 1), with individual contributions of 53.5, 25.6, 14.3 and 6.1% for each dimension, respectively. For ease of visualisation, the two-dimensional representation is shown in Figure 3 and the obtained Euclidean distances in the four-dimensional space are shown in Figure 4 together with the results of the instrument-in-noise tests.

The results in Figure 3 suggest that the pianos (so far, limited to the note C#$_5$) can be classified into four distinct groups: pianos 1-5, pianos 2-7, pianos 3-6 and piano 4. Although piano 4 seems to have a middle similarity with all these groups, in the four-dimensional space its dis-

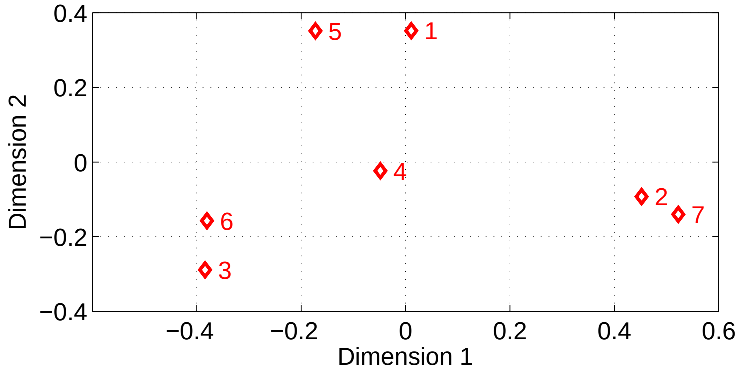


Figure 3: Representation of the similarity matrix in a two-dimensional perceptual space (goodness-of-fit of 79.1%). These results suggest that the recordings of note C#$_5$ can be classified in four groups: pianos 1-5, 2-7, 3-6 and piano 4.

tances increase systematically. The distances for all the other pianos do not differ considerably with respect to the ones in the two-dimensional representation.

4.2 Instrument-in-noise identification

The results for the instrument-in-noise experiments are shown in Figure 4. On average there were 6 estimates per piano pair (minimum of 4 and maximum of 10). Thresholds above 20 dB were removed, since for those conditions the noise is almost inaudible, meaning that subjects were not able to distinguish the piano sounds even in silence. This corresponds to a violation of one assumption of the staircase procedure: recognition at easy conditions should result in scores of about 100%. In total only 6 (out of 131 estimates) were removed, out of which 3 estimates occurred systematically in the comparison between piano 3 and 6 and the other three occurred once when evaluating different pairs.

4.3 Comparison between methods

To objectively measure how similar the performance of the two methods is, the thresholds of the instrument-in-noise task were correlated with the Euclidean distances of the triadic comparisons. Excluding the data points corresponding to the piano pair 3-6, where no thresholds within the resolution of the instrument-in-noise test were obtained, a significant normalised (Pearson) cross covariance of -0.48 was found. Although both methods are not perfectly consistent, a common tendency can be observed. Two exceptions are the judgements for the piano pairs 2-7 and 1-5. Both pairs are judged as being more (acoustically) similar by the triadic comparison method. Despite these differences, we believe that this instrument-in-noise method is a promising method to quantify acoustic differences between sounds, which is correlated with subjective similarity judgements.

5 Conclusion

In this paper we have presented a method to compare piano sounds, measuring their acoustic similarity based on an instrument-in-noise test, where the noise is matched to the spectro-

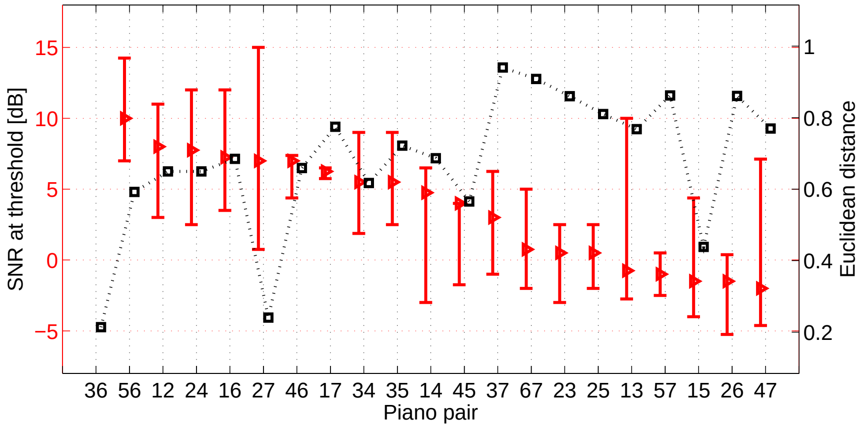


Figure 4: Identification thresholds for the instrument-in-noise tests (red triangle markers) and Euclidean distances from the triadic comparisons (black squared markers). The piano pairs are shown along the abscissa and were ordered from most to least similar (higher to lower thresholds in SNR) using the median across participants as estimate of similarity. The error bars represent interquartile ranges. The Euclidean distances range between 0.21 (pair 3-6) and 0.94 (pair 3-7). Lower and higher distances represent higher and lower similarity, respectively. For a perfect consistency between methods, the triadic comparison results should increase monotonically.

temporal properties of the test sounds. The results were compared to the method of triadic comparisons, which is a widely used method to map sounds into a perceptual similarity space. Initial results when comparing the note C$\#_5$ played in 7 old Viennese pianos with 14 subjects had a normalised cross covariance of -0.48. This value denotes a moderate inverse correlation between thresholds and Euclidean distances, meaning that the higher the threshold should result in a lower Euclidean distance. The next step in our research is to correlate the acoustic similarity with the properties of the test piano sounds. As mentioned earlier, the pianos we used in this study have important construction differences in, e.g., their hammers, strings and soundboards. These differences affect the acoustic properties of the sounds as, the envelope, spectral content and attack and decay time [2, 5]. Our instrument-in-noise method is a tool that detriments the acoustic properties of the test sounds, allowing us to evaluate which cues are more prominent at the participant's discriminability thresholds. With this information, a hierarchical ordering of these cues can be established (ongoing work) which will be valid, in our case, for notes in the higher frequency range of the piano (around C$\#_5$). Based on the results presented in this paper, we believe that this instrument-in-noise method is a promising method to quantify acoustic differences between sounds. The method can be used to perform any other within-instrument comparison as long as the test sounds have similar fundamental frequencies and duration.

Acknowledgements

This research work has been funded by the European Commission within the ITN Marie Curie Action project BATWOMAN under the 7th Framework Programme (EC grant agreement N^{o} 605867) and has been further supported by the Lise-Meitner Fellowship M1653-N30 of the Austrian Science Fund (FWF) attributed to Antoine Chaigne.

References

[1] N. Giordano. "The invention and evolution of the piano". *Acoustics Today* 12, (2016), pp. 12–19.

[2] A. Chaigne, J. Chabassier, and M. Duruflé. "Energy analysis of structural changes in pianos". *Proceedings of the Third Vienna Talk on Music Acoustics*. 2015, pp. 184–191.

[3] J. Chabassier, A. Chaigne, and P. Joly. "Time domain simulation of a piano. Part 1: model description". *ESAIM: Mathematical Modelling and Numerical Analysis* 48, (2013), pp. 1241–78.

[4] J. Chabassier, M. Durufle, and P. Joly. "Time Domain Simulation of a Piano. Part 2: Numerical Aspects". *ESAIM: Mathematical Modelling and Numerical Analysis*, (2014), pp. 1–43.

[5] A. Chaigne, M. Hennet, J. Chabassier, and M. Duruflé. "Comparison between three different Viennese pianos of the nineteenth century". *International Congress on Acoustics*. Buenos Aires, 2016.

[6] W. J. Levelt, J. P. van de Geer, and R Plomp. "Triadic comparisons of musical intervals". *Br J Math Stat Psychol* 19, (1966), pp. 163–179.

[7] C. Fritz, J. Woodhouse, F. P.-H. Cheng, I. Cross, A. F. Blackwell, and B. C. J. Moore. "Perceptual studies of violin body damping and vibrato". *J Acoust Soc Am* 127, (2010), pp. 513–524.

[8] A. Novello, M. McKinney, and A. Kohlrausch. "Perceptual evaluation of inter-song similarity in Western popular music". *Journal of New Music Research* 40, (2011), pp. 1–26.

[9] W. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann. "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment". *Int J Audiol* 40, (2001), pp. 148–157.

[10] M. Schroeder. "Reference signal for signal quality studies". *J Acoust Soc Am* 44, (1968), pp. 1735–1736.

[11] B. De Man and J. Reiss. "APE: Audio Perceptual Evaluation Toolbox for MATLAB". *Audio Engineering Society Convention 136*, (2014), pp. 1–4.

[12] T. Francart, A. van Wieringen, and J. Wouters. "APEX 3: a multi-purpose test platform for auditory psychophysical experiments." *Journal of neuroscience methods* 172, (2008), pp. 283–93.

[13] B. Glasberg and B. C. J. Moore. "A model of loudness applicable to time-varying sounds". *J Audio Eng Soc* 50, (2002), pp. 331–342.

[14] B. Everitt. "Multidimensional scaling and correspondence analysis". *An R and S-PLUS companion to multivariate analysis*. Springer Verlag, 2005. Chap. 5, pp. 91–114.