Perceptual evaluation of differences between original and synthesised musical instrument sounds: the role of room acoustics

Alejandro Osses V.[1)], Chungeun Kim[1)] and Armin Kohlrausch[1,2)]
[1)] Human-Technology Interaction group, Department of Industrial Engineering & Innovation Science, Eindhoven University of Technology, Eindhoven, the Netherlands.
[2)] Behavior, Cognition & Perception group, Philips Research Europe, Eindhoven, the Netherlands.

Summary
To better understand the properties of musical instruments, their recordings in controlled situations are often compared with computational models that recreate such situations. The purpose of this study was to apply psychoacoustic descriptors to measured and modelled signals of a musical instrument and to identify whether there were perceptual differences in the signals. The hummer was chosen, which is a musical instrument that generates sound when it is rotated at certain speeds, due to airflow exciting the resonance frequencies of the system. In a previous study, we found perceptual differences for loudness fluctuations and roughness for an evaluation in anechoic conditions. In this study, we have applied similar evaluation criteria to audio signals that account for the first reflections coming from one room surface. Perceptual differences were found in the perceived pitch when adding the reflected sounds into the computational model. For one rotation condition, differences in loudness were also found.

PACS no. 43.66.Cb, 43.75.Cd, 43.75.Zz, 43.55.Br

1. Introduction

Characterising musical instruments has always been challenging, since each musical instrument has different properties. Among others, there are differences in terms of the intensity and frequency range covered (higher or lower pitch sounds, tonal or atonal sound radiation, etc.). To better understand their properties, sound recordings in controlled situations (original sounds) are often compared with computational models (synthesised sounds) that recreate such situations. However, when comparing measured and modelled audio signals many researchers make use of descriptors similar to those used in physical applications, electronics or telecommunications, where human perception is not necessarily involved.

Even considering that the human hearing system is not yet fully understood, psychoacoustic studies have addressed the problem of sound perception by developing and validating audio descriptors based on perceptual criteria. Psychoacoustic audio descriptors have mostly been developed using artificial stimuli [1, 2, 3] and less often using speech, music or other everyday sounds [4, 5, 6]. In a previous study [6], we performed a perceptual evaluation of an instrument called hummer. For this instrument, synthesised sounds through computational modelling were already available, as well as the measured original sounds in anechoic and semi-reverberant conditions [7, 8]. In that study we compared the original and synthesised sounds in the anechoic conditions, finding differences in loudness fluctuations and roughness. The purpose of this study is to apply similar methods to the signals in the semi-reverberant condition to evaluate (1) whether the same analysis procedures are applicable and (2) whether the differences found previously are still present. First an overview of the basic principles of the hummer is given, followed by a description of the parameters of fundamental frequency and loudness and the results of their use in the evaluation. Finally, the applicability of these methods to other musical instruments are discussed.

2. Characteristics of the hummer

The hummer is a flexible plastic corrugated pipe with both ends open. A schematic geometry of the hummer and typical dimensions are shown in Figure 1.

Copyright© (2015) by EAA-NAG-ABAV, ISSN 2226-5147
All rights reserved

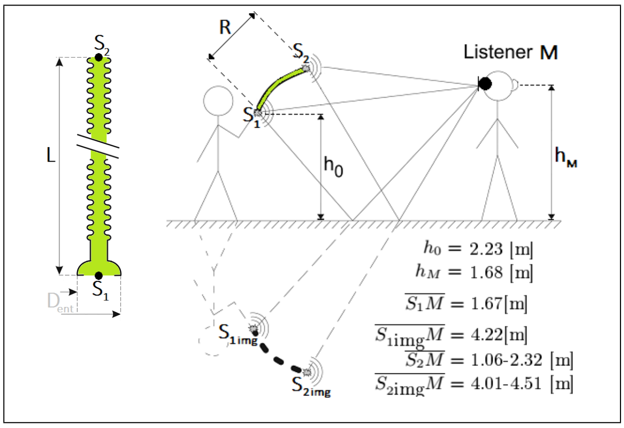

Figure 1. Schematic drawing of a hummer. In this study, L =70 cm, the inlet (S_1) had an entrance diameter D_{ent} of 3.3 cm. The opposite end of the hummer was identified as the outlet (S_2).

To generate sound, the hummer has to be rotated at a certain speed in order to excite the natural frequencies of the pipe. The resonance frequencies f_n of the system as a function of the acoustic mode n are given by:

$$f_n \approx n\frac{c_{\text{eff}}}{2L} \qquad \text{with } n = 2, 3, ... \tag{1}$$

where c_{eff} corresponds to the effective speed of sound in the tube and L corresponds to the length of the pipe. The estimated effective speed of sound is approximately 310 m/s [7]. Resonance frequencies estimated using (1) and those obtained from measurements are presented in Table I.

Table I. Resonance frequencies f_n and rotation period Ω_n for the hummer at different rotation speeds (modes 2 and 4) considering both theory (1) and measurements.

Acoustic mode n	Frequency f_n [Hz] Theory	Frequency f_n [Hz] Measured	Δ F0 [%]	Period Ω_n [s]
2	442.9	424.4	4.2	0.6023
4	885.7	851.8	3.8	0.2961

The rotational movement of the hummer produces a periodic variation in distance between sound source and listener, which leads to positive and negative frequency shifts due to the Doppler effect. This variation is related to the rotation period of the hummer.

3. Methods

The measured and modelled data of the hummer instrument were provided by A. Hirschberg and are described in the following subsections. A more extensive description of the measurement set-up can be found in [8].

3.1. Original sounds

The set-up was installed in a semi-anechoic room with a reflecting floor and a volume of 100 m^3. Recordings of the hummer were made at different rotation speeds in two conditions: reflecting and non-reflecting floor. An absorptive mat was placed between the vertical support rod of the hummer and the location of the listener (microphone) to obtain the non-reflective condition, i.e., considering only the contributions from sources S_1 and S_2. The reflecting condition considered the contribution of sources S_1, S_2 and also from the imaginary sources $S_{1\text{img}}$ and $S_{2\text{img}}$ (see Figure 1). The rotation speeds were controlled by an electrical motor.

The hummer was attached to the spikes of a 26" bicycle wheel. The inlet S_1 was placed close to the axis of rotation (wheel axis). The outlet S_2 was at a distance of 0.70 m from the wheel axis, approximately 0.30 m outside the radius of the wheel. The wheel was mounted on a structure (oriented horizontally), at a height of 2.23 m above the floor. The wheel axis was defined to be at coordinates (0,0,2.23) m.

Two microphones B&K type 4190 were used to record the hummer. In this study, only the microphone located at coordinates (1.58, 0, 1.68) m was used. This coordinate corresponds to a distance of 1.67 m from the centre of rotation (wheel axis). Each recording had a duration of 20 s and was sampled at 10 kHz, with an amplitude resolution of 16 bits. The measured resonance frequencies differed by about 4% from the approximation given by (1), as shown in Table I.

The measured signals were re-sampled at 44.1 kHz, with an amplitude resolution of 16 bits. All results obtained for frequencies above the Nyquist frequency of 5 kHz were ignored (corresponding to about 20 Bark). The average level was adjusted according to the reference levels of 60 and 78 dB SPL at 0.85 m from the origin of the system for the acoustic modes 2 and 4, respectively.

3.2. Synthesised sounds

Considering a hummer of length L = 0.7 m, as represented in Figure 1, the instrument can be modelled as two monopole sound sources. The inlet, near the axis of the wheel and having an entrance diameter of D_{ent} = 3.3 cm, was modelled as a fixed source S_1, while the outlet was modelled as a rotating source S_2 with a rotation period of Ω_n. Because of the flexible nature of the hummer, an effective rotation radius R of 0.67 m was used.

The synthesised waveforms were obtained using the model developed in [7, 8], which accepts L, D_{ent}, R,

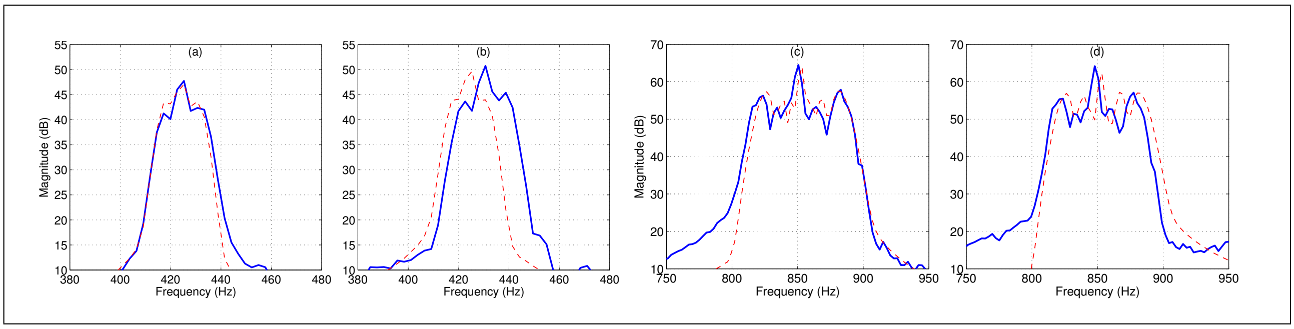

Figure 2. Frequency spectra of measured (continuous) and modelled (dashed) hummer signals at the listener's position: in acoustic mode 2, (a) non-reflecting and (b) reflecting conditions; and in mode 4, (c) non-reflecting and (d) reflecting conditions.

Ω_n, f_n and the parametrised positions of the sound sources $S_{1,2}(t)$ and the listener (microphone location) as input parameters. The measured resonance frequencies f_n and rotation periods Ω_n presented in Table I were used instead of the respective theoretical values.

The modelled signals were sampled at 44.1 kHz with an amplitude resolution of 16 bits. The average level was adjusted according to the reference levels of 60 and 78 dB SPL at 0.85 m from the origin of the system for the acoustic modes 2 and 4, respectively.

3.3. Evaluation criteria

The measured signals of the hummer cover relatively wider frequency ranges outside the resonance frequencies f_n (with $n = 2, 4$), whereas the synthesised sounds show spectra more concentrated around the resonance frequency, as shown in Figure 2. To reduce the differences between the measured and modelled signals only a frequency range around the frequencies f_n was evaluated. In our previous study, we found that those differences were related to noise, which was not taken into account in the computational model of the hummer. These frequency limits were labelled as $z_{\text{min, max}}$ and are shown in Table II. We have chosen psychoacoustic descriptors that account for level perception (loudness) using a model that accounts for spectral and to some extent temporal masking patterns to investigate whether there are significant perceptual differences between the compared signals. Furthermore, fundamental frequency contours, which correspond to an estimate of the pitch periodicity of the sounds, were used to evaluate their variation in frequency shifts.

Fundamental frequency F0: fundamental frequency estimates were obtained using the Praat software [9, 10]. In this study, F0 estimates were used to investigate frequency variations (Doppler shift) of the hummer sounds, particularly the difference between the minimum and maximum estimates. For sinusoidal frequency-modulated sounds ($f_{\text{mod}} = 4$ Hz) varying in $\pm\Delta f$ around a carrier frequency f, the just-noticeable changes in frequency of 0.42% ($f_2 = 424.4$ Hz) and 0.35% ($f_4 = 851.8$ Hz) are obtained for acoustic modes 2 and 4, respectively [11].

Loudness: loudness was evaluated using the Dynamic Loudness Model DLM [12] which provides loudness excitation patterns in time and in frequency. In this paper we used the model outputs of main loudness and critical band levels L_G related to the maximum and minimum specific loudness patterns. When obtaining L_G levels by applying an inverse transformation of the specific loudness patterns, both spectral and temporal resolution of the human ear are accounted for [12]. Regarding the latter aspect, we used the L_G levels as an estimate of the forward masking patterns of the hummer signals.

4. Results

The results of the fundamental frequency estimation and the perceptual analysis using loudness for the hummer signals are presented in this section.

4.1. Fundamental frequency

The results for the fundamental frequency estimation are shown in Figure 3, where a pitch estimate was found for every audio segment (length of 100 ms, hop-size of 15 ms, F0 candidates between 30 and 1400 Hz). The differences between the measured and modelled F0 estimates (normalised to f_n) were found to exceed the just-noticeable variations in frequency reported in the literature for stationary frequency-modulated tones, with maximum absolute values ranging from 3 to 5%. If we consider that the evaluated sounds have modulations with dynamic variation (Doppler effect) rather than stationary modulations, then these differences would be unlikely to be perceived. However, the modelled signals in the reflecting condition of mode 2 and in the non-reflective condition of mode 4 showed a decrease in the frequency deviation ($F0_{\text{max}} - F0_{\text{min}}$)

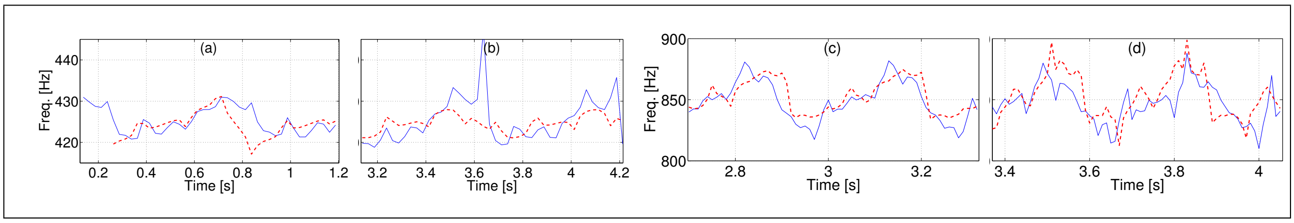

Figure 3. Fundamental frequency F0 estimation for measured (continuous) and modelled (dashed) hummer signals: in mode 2, (a) non-reflecting and (b) reflecting conditions; and in mode 4, (c) non-reflecting and (d) reflecting conditions.

compared to the measured signals, as shown in Figure 3, panels b and c, respectively.

4.2. Loudness

The results for main loudness and critical band levels L_G are presented in Figures 4, 5 and 6. The percentiles L_5, L_{50} and L_{95} calculated from the specific loudness patterns are summarised in Table II. The results were obtained by averaging two rotation periods of the hummer signals. For the original sounds the most stable periods were chosen. When ignoring the contribution of the frequency components outside the range defined by z_{min} and z_{max}, the loudness differences ΔL_{50} in acoustic mode 2 were 0.1 and 0.4 sone for the non-reflecting and reflecting conditions, respectively. Likewise, the differences in mode 4 were 0.3 and 0.5 sone. A positive difference means that the measured signal had a larger value. In mode 2, the minimum L_5 and maximum L_{95} values had a good agreement with a deviation of 0.1 sone in both non-reflecting and reflecting conditions. In mode 4, the modelled signals had lower minimum values (ΔL_5 of 0.8 and 0.5 sone), while the maximum values were larger (ΔL_{95} of -0.8 and -0.7 sone). These differences were also examined using the critical band levels L_G, which can be used as estimators of the minimum and maximum masking thresholds. In mode 4, the differences were larger in the minimum masking patterns with values from 3.6 to 4.8 dB below the masking values found for the measured signals, as shown in panels (g) and (h) of Figure 6. On average, the masking thresholds for the other conditions had differences mostly below 1 dB as shown in panels (e) and (f) of Figure 5.

5. Discussion

In this section the differences found between the original and synthesised sounds are discussed, followed by the applicability of the psychoacoustic descriptors to other musical instruments.

5.1. Differences between the original and synthesised sounds

The reduction of the Doppler effect ($F0_{\text{max}} - F0_{\text{min}}$) found in the reflecting condition of acoustic mode 2, as shown in panel (b) of Figure 3, is a consequence of the interference between the direct sounds of the sources S_1 and S_2 and the reflected sounds. In acoustic mode 4, only considering the direct sounds, the frequency deviation was $2 \cdot \Delta f = 35$ Hz (panel c) while for the reflecting condition this deviation increased to 60 Hz (panel d). In this case the interferences emphasised the Doppler effect at the listener's position. This suggests that the Doppler effect in the computational model is strongly influenced from the phase of the sources S_1 and S_2. The model considers that both sources are radiating sounds either in phase (for even modes) or out of phase (for odd modes).

When integrating the main loudness values using a reduced frequency range around the resonance frequencies, as presented in Table II, the periodicity of the measured signals becomes clearer, especially in mode 2, where due to the slower rotations larger variations in the hummer levels were present. The minimum and maximum masking patterns for both, the non-reflecting and reflecting conditions had close values in mode 2. In mode 4, the differences between measured and modelled signals had, for both conditions, a good agreement for the maximum masking patterns but differences between 3.6 to 4.8 dB for the minimum patterns, as seen in panels (g) and (h) of Figure 6. This means that the reflected sounds do not have a large influence on the masking patterns. An increase in the minimum masking patterns can be reached by increasing the contribution of the non-fluctuating sound source, i.e., the stationary source S_1. In the computational model, both sources S_1 and S_2 are considered to radiate sound at the same level. Some additional calculations with various levels of S_1 showed that an increase of 2 dB will reduce the difference in the minimum masking pattern by 1-2 dB.

In summary, we have found differences in the psychoacoustic descriptors between measured and modelled hummer signals when comparing them in reflecting and non-reflecting conditions. Overall the differences between conditions were clearer for the modelled signals and they can be reduced by introducing (1) a gain factor for the levels radiated from the moving and the stationary source; (2) a starting phase for the radiation of the sources, or (3) an absorption

Table II. Summary of the specific loudness patterns in percentiles for 2 periods of rotation of the hummer signals. Percentile 5 and 95 represent minimum and maximum values, respectively. Percentile 50 is an estimate of the mean loudness value. To assess these values, only the frequency components in the range ($z_{\mathrm{min}}, z_{\mathrm{max}}$) were taken into account.

Acoustic Mode n / Type	Frequency limits [Bark] z_{min}-z_{max}	Loudness [sones] Non-reflecting condition L_5	L_{50}	L_{95}	$L_{95}-L_5$	Reflecting condition L_5	L_{50}	L_{95}	$L_{95}-L_5$
2 / measured	2.9 - 8.5	1.1	2.1	2.7	1.6	1.0	2.3	3.0	2.0
2 / modelled	2.9 - 8.5	0.9	2.0	2.6	1.7	0.9	1.9	2.9	2.0
4 / measured	6.0 - 10.7	3.9	5.7	7.0	3.1	3.6	5.5	6.8	3.2
4 / modelled	6.0 - 10.7	3.1	5.4	7.8	4.7	3.1	5.0	7.6	4.5

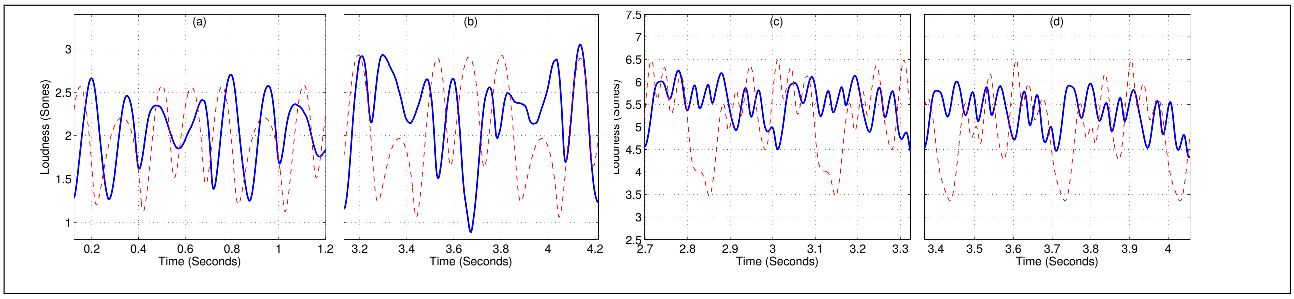

Figure 4. Loudness of measured (continuous) and modelled (dashed) hummer signals: in mode 2, (a) non-reflecting and (b) reflecting conditions; and in mode 4, (c) non-reflecting and (d) reflecting conditions. Only the loudness contribution of frequency components between z_{min} and z_{max} were taken into account.

coefficient different from 1 or 0 (non-reflecting and reflecting conditions respectively).

5.2. Applicability in the evaluation of musical instruments

The methods presented in our previous hummer study were based on stationary (harmonic) sounds evaluated in anechoic conditions. In this study we have broadened the evaluation condition from anechoic (non-reflecting) to a simple reflecting condition (reflections from one surface). The forward masking phenomenon was, to some extent, taken into account by using the critical band levels available from the DLM model summed to the spectral integration also accounted in this model. With this approach, it was possible to point out some perception-related differences between the measured and modelled sounds towards more realistic sound synthesis. However, some aspects such as the effect of temporal transients have not been evaluated, making these methods not immediately extensible to musical instruments with a strong time-varying component.

6. Conclusions

A perceptual evaluation of original and synthesised sounds of the hummer based on the loudness and fundamental frequency estimates was presented. Considering the outcomes of a previous study –where we found perceptual differences mainly in high frequencies– we focused here on psychoacoustic measures taken from the frequencies containing the resonance frequencies of acoustic modes 2 and 4. Moving the source further away (from 0.85 m to 1.67 m) and adding a condition with reflections from one surface, the modelled signals were found to be more influenced by the constructive and destructive interferences than the measured signals. Some aspects that have not been addressed in this study are: (1) the influence of temporal transients; (2) study of non-harmonic sounds. For instance, to approach the problem of temporal transients time-varying masking models could be used [13]. A more elaborated approach to room acoustics could be taken by using, e.g., auralisation techniques. However, we believe that the analysis presented in this study might provide more relevant psychoacoustic results when comparing modelled and measured signals of a musical instrument than the use of objective measures as sound pressure level or spectrograms.

Acknowledgement

We are grateful to Oleksii Rudenko, Michel Hirschberg and Prof. Avraham Hirschberg for providing the recordings, the computational model of the hummer and the time spent to explain us the details of the data. 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 no. 605867).

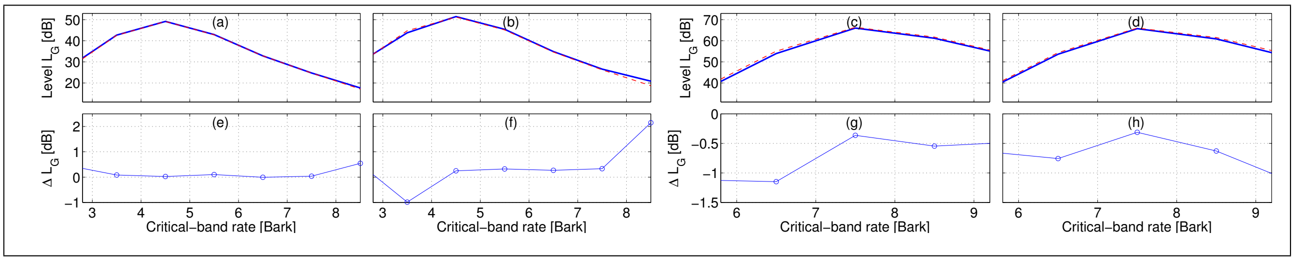

Figure 5. Maximum critical band levels L_G [dB] for measured and modelled hummer signals: in acoustic mode 2, (a) non-reflecting and (b) reflecting conditions; and in mode 4, (c) non-reflecting and (d) reflecting conditions. Below each panel, the differences between the measured and modelled signals are shown. The differences were overall below 1 dB. Because of this, the L_G levels in the top panels seem to be overlapped, however they are still shown for ease of comparison of these results with the top panels of Figure 6.

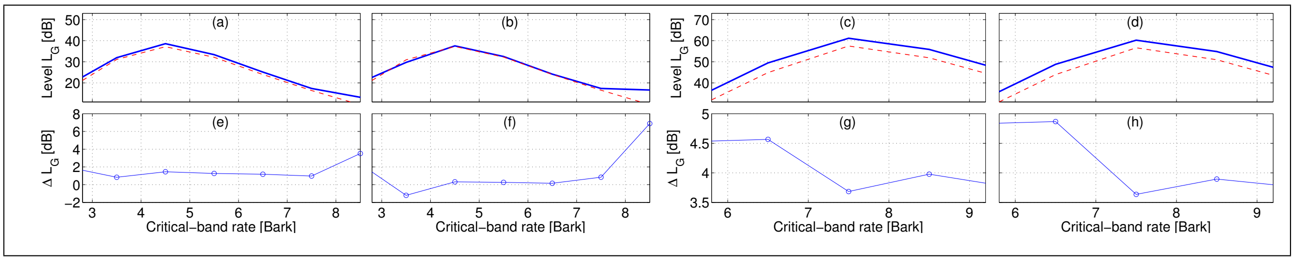

Figure 6. Minimum critical band levels L_G [dB] for measured (continuous) and modelled (dashed) hummer signals: in acoustic mode 2, (a) non-reflecting and (b) reflecting conditions; and in mode 4, (c) non-reflecting and (d) reflecting conditions. Below each panel, the differences between the measured and modelled signals are shown. The maximum difference was 4.9 [dB] at 6.5 Bark as shown in panel (h).

References

[1] Zwicker, E. (1977). "Procedure for calculating loudness of temporally variable sounds". *J Acoust Soc Am* 62.(3), pp. 675–682.

[2] Kemp, S. (1982). "Roughness of frequency-modulated tones". *Acustica* 50, pp. 126–133.

[3] Daniel, P. and R. Weber (1997). "Psychoacoustical roughness: implementation of an optimized model". *Acustica* 83, pp. 113–123.

[4] Terhardt, E. (1978). "Psychoacoustic evaluation of musical sounds". *Perception & Psychophysics* 23.(6), pp. 483–92.

[5] Aramaki, M., H. Bailleres, L. Brancheriau, R. Kronland-Martinet, and S. Ystad (2007). "Sound quality assessment of wood for xylophone bars". *J Acoust Soc Am* 121.(4), p. 2407.

[6] Osses, A and A Kohlrausch (2014). "Perceptual evaluation of differences between original and synthesised musical instrument sounds". *IX Congreso Iberoamericano de Acústica FIA 2014*. Ed. by J. Arenas. Valdivia, Chile, pp. 987–97.

[7] Nakiboğlu, G., O. Rudenko, and A. Hirschberg (2012). "Aeroacoustics of the swinging corrugated tube: voice of the Dragon." *J Acoust Soc Am* 131.(1), pp. 749–65.

[8] Hirschberg, M., O. Rudenko, G. Nakiboğlu, A. Holten, J. Willems, and A. Hirschberg (2013). "The voice of the mechanical dragon". *Proceedings of SMAC*. Stockholm.

[9] Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". *Proceedings of the Institute of Phonetic Sciences* 17, pp. 97–110.

[10] Boersma, P. and D. Weenink (2001). "Praat, a system for doing phonetics by computer". *Glot Int.* 5.(9/10), pp. 341–345.

[11] Fastl, H. and E. Zwicker (2007). *Psychoacoustics, Facts and Models*. Third edit. Springer Berlin Heidelberg.

[12] Chalupper, J. and H. Fastl (2002). "Dynamic loudness model (DLM) for normal and hearing-impaired listeners". *Acta Acustica united with Acustica* 88.(3), pp. 378–386.

[13] Dau, T., D. Püschel, and A. Kohlrausch (1996). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure." *J Acoust Soc Am* 99.(6), pp. 3615–22.