Audio Engineering Society

Convention Paper 8367

Presented at the 130th Convention
2011 May 13–16 London, UK

The Effect of Loudness Overflow on Equal-Loudness-Level Contours

ANDREW J.R. SIMPSON AND JOSHUA D. REISS

Centre for Digital Music, Queen Mary University of London, London E1 4NS, UK
andy.simpson@elec.qmul.ac.uk
josh.reiss@elec.qmul.ac.uk

ABSTRACT

This paper presents a formal derivation of the Loudness Overflow Effect (LOE), which describes the impact of nonlinear distortion on loudness. Computational analysis is then performed, comprised of two experiments involving two compressive static nonlinearities, and using two well-known time-varying loudness models. The results characterize the nonlinearities in terms of LOE as a function of frequency and of listening level in the case of 250-ms pure-tone stimuli, and in terms of the traditional equal-loudness-level contours. The analysis is then extended to synthesized wind instruments for one of the nonlinearities. The effect of the nonlinearity on loudness as a function of musical note fundamental frequency and listening level is described for various synthesized instruments.

1. INTRODUCTION

The present work analyses the effects of compressive static nonlinear distortion on the loudness function in the case of pure-tone stimuli, as used to derive the traditional equal-loudness-level contours, *e.g.*, [1]. In a previous study [2], the Loudness Overflow Effect (LOE) was introduced to describe the effects of nonlinear distortion on loudness. This study seeks to establish a framework for the characterization and investigation of audio system nonlinearities in terms of signal-dependent LOE behavior.

A formal framework is given to describe LOE and followed by the results of a computational analysis based on the loudness models of Zwicker and Fastl [3] and Glasberg and Moore [4]. Initially, two experiments are presented and used to characterize a pair of compressive static nonlinearities. The loudness models are used to predict LOE behavior as a function of frequency and of listening-level. The loudness models are then used to produce the traditional equal-loudness-level contours. In each case, results describe deviations in the loudness function caused by nonlinearity. Next, two further experiments extend this analysis from the artificial stimulus of the pure-tone signal to the more realistic stimulus of the synthesized musical wind-instrument note. Finally, the loudness model of [4] is used to predict LOE behavior as a function of musical-note frequency and listening-level, and to produce equal-loudness-level contours for the musical notes. The loudness model is also used to produce a time-varying LOE function for monophonic synthesized music signals at various listening levels. Brief discussion of the psychoacoustics of LOE and of the two models is given.

2. LOUDNESS OVERFLOW EFFECT

The compressive type nonlinearity is common to audio systems [5, 6], and there is generally an expectation that a compressive nonlinearity will result in reduced loudness, as shown in [7]. Yet distortion products introduced by the nonlinearity may compensate for the amplitude compression in the loudness function [2]. In generalized loudness gain terms, LOE can be formalized as follows. Where the time-varying loudness (L), of the time-domain signal amplitude (x), at time (t) is denoted $L(x,t)$ and where the nonlinear input-output function can be described with a function $f(x)$, LOE is defined as:

$$LOE(f,t) = L(f(x),t) - L(x,t) \qquad (1)$$

Positive values of LOE indicate an increase in loudness and negative values indicate a decrease in loudness as a result of the nonlinearity.

3. PURE-TONE LOE(f, ℓ)

The loudness models of Zwicker and Fastl [3] and Glasberg and Moore [4] were used to produce a normalized LOE function of frequency and of listening level, using 250-ms duration pure-tone sinusoidal and nonlinearly distorted sinusoidal signals as stimulus. Each loudness model was input with pure-tone sinusoidal signals at frequencies between 20 – 10,000 Hz, at simulated peak levels between 40 – 120 dB SPL. A compressive nonlinearity was then applied to each signal, the resulting signal input to the model. Difference in maximum short-term loudness was calculated according to Eq. 1 for each pair of signals and plotted as LOE graphs as a function of frequency and as a function of listening-level.

3.1. Characterization of the Nonlinear Functions

Two typical nonlinear functions were implemented according to the following expressions:

$$f(x) = 0.5(x + |x|^{1.02} - x^2) \qquad (2)$$

$$f(x) = \frac{x}{|x|}\begin{cases} 2|x| & for \quad 0 \leq |x| \leq 1/3 \\ \frac{3-(2-3|x|)^2}{3} & for \quad 1/3 \leq |x| \leq 2/3 \\ 1 & for \quad 2/3 \leq |x| \leq 1 \end{cases} \qquad (3)$$

Fig. 1(a) shows the normalized input-output functions of Eq. 2 and 3 respectively. Fig. 1(b) shows the peak compression effects of Eq. 2 and 3. Figs. 1(c) and 1(d) show the power spectral density functions of a 1 kHz pure-tone (at unity) processed with Eq. 2 and 3 respectively. Eq. 2 can be characterized as a smooth, gradual-onset, low-order, symmetrical, saturating nonlinearity with a 1-kHz THD+N value of ~14% and an equivalent peak compression of ~6 dB (Fig. 1(b)). Eq. 3, taken from Schetzen [8], can be characterized as a soft-clipping function featuring a discontinuity of slope and is of less-gradual onset, low-order, symmetrical, saturating nonlinearity with a 1-kHz THD+N value of ~22% and an equivalent peak compression of ~6 dB (Fig. 1(b)). The 1-kHz power spectral density functions (Fig. 1(c) and 1(d)) feature odd-order harmonic distortion products which are the result of symmetry in the nonlinearities.

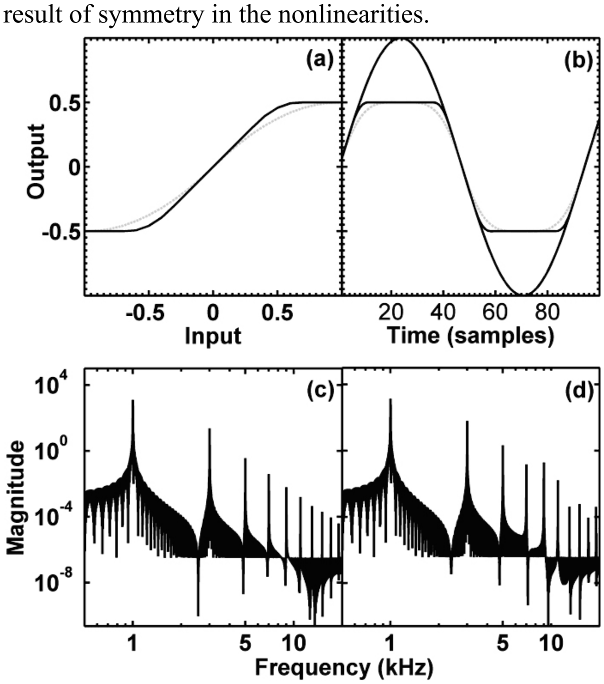

Figure 1. (a) Normalized input-output functions of Eq. 2 (*black*) and Eq. 3 (*grey*). (b) Peak compression effects of Eq. 2 (*black*) and Eq. 3 (*grey*) on a sinusoidal signal. (c, d) Power spectral density, 1 kHz as processed with Eq. 2 (*left, lower*) and Eq. 3 (*right, lower*).

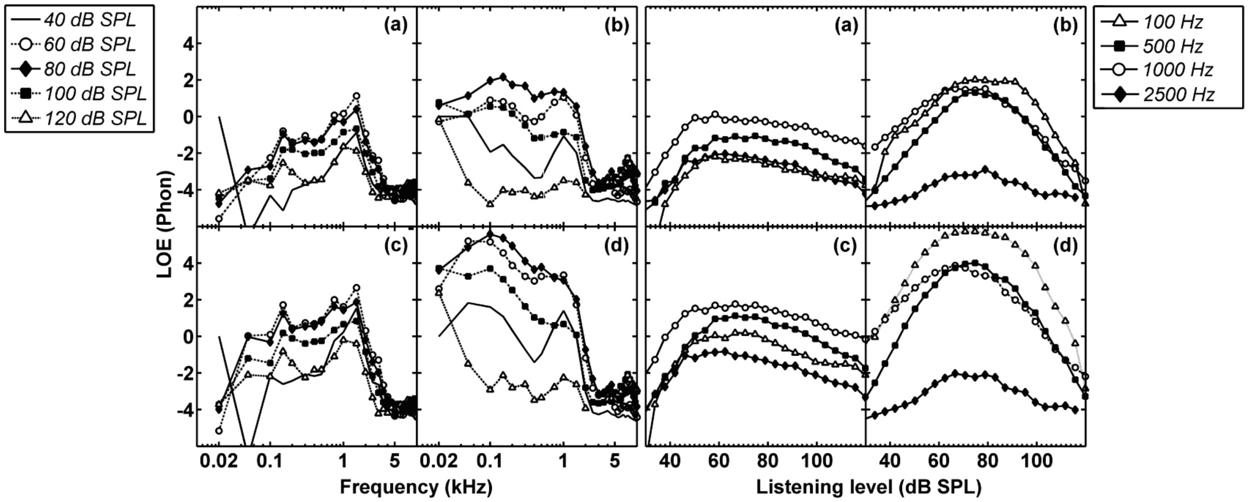

Figure 2 (*left panel*) LOE as a function of frequency. Figure 3 (*right panel*) LOE as a function of listening-level. 250-ms pure-tone signals. Calculated from results taken from the models of [3] (*a, c*) and [4] (*b, d*). Compressed using Eq. 2 (*a, b*) and Eq. 3 (*c, d*).

3.2. Stimuli

The loudness model was operated at 44.1 kHz and 250-ms pure-tone sinusoidal signals were generated digitally at the same rate. The pure-tone signals were then 100 times oversampled, nonlinearly processed *at unity* and decimated to the original sampling rate with an eighth-order anti-alias filter. This prevented the introduction of aliased (under-sampled) ultrasonic harmonic distortion products into the sonic range.

3.3. Results and Discussion

Figures 2 and 3 show LOE as a function of frequency and of listening-level, for Eq. 2 and Eq. 3, and for the models of [3] and [4]. Some features are consistent for both models and both nonlinearities. The results feature a strong dependence on frequency and listening-level. At listening levels between 60 – 100 dB SPL the largest positive LOE values are seen. At high frequencies (>2 kHz) harmonic distortion products contribute little to the loudness sum and so the result is a reduction in loudness (negative LOE values). At lower frequencies (< 2 kHz), an increase in loudness is shown (positive LOE values). For both nonlinearities and both loudness models, a peak compression of ~6 dB (Fig. 1(b)) has resulted in a net *increase* of loudness level at low frequencies and at medium listening levels. In the most extreme case, for the model of [4] the peak compression of ~6dB has resulted in an increase in loudness level of ~6 dB (Fig. 3(d)). While the peak compression produced by both nonlinear functions is almost identical at ~6dB, the resulting LOE functions are quite different. The greater magnitude of harmonic distortion products of Eq. 3 produce predictably greater effects on the LOE function than those shown as a result of Eq. 2.

There are two key differences between the models. First, [3] applies the Bark scale and [4] applies the equivalent rectangular bandwidth (ERB) scale. Second, the model of [4] is designed to account for partially-masked or *sub-threshold* loudness, where multiple sub-threshold components are able to sum to audibility. The Bark scale is derived by the classical masking experiment method involving a narrowband masker and probe tone [9], and has since been superseded by the ERB scale applied in the model of [4], which is derived according to the notched-noise masker method [10]. The bandwidths of the Bark scale are significantly wider than that of the ERB [11] and therefore predict greater masking of the harmonic distortion products. In combination with sub-threshold partial-masking behavior, it appears plausible that these two differences may account for the difference in results.

4. EQUAL-LOUDNESS-LEVEL CONTOURS

A recursive algorithm was used to simulate equal-loudness-levels similar to the method-of-adjustment task employed in traditional determination of equal-loudness-level contours [12]. For each sound pressure level, the loudness level (phon) of a 250-ms 1-kHz tone was calculated and 250-ms tones at various other frequencies were adjusted in sound pressure level until their simulated loudness level was equal to that of the 1 kHz tone. The algorithm recursively adjusted the sound pressure level of the target tone by a value inversely proportional to the difference in predicted loudness level of the reference 1-kHz tone until convergence to within 1 dB (phon). The tones were then processed by the nonlinear function (Eq. 2 and 3) and the process was repeated. Pure-tone signals were produced digitally at a sample rate of 44.1 kHz and processed using the over-sampled method described previously.

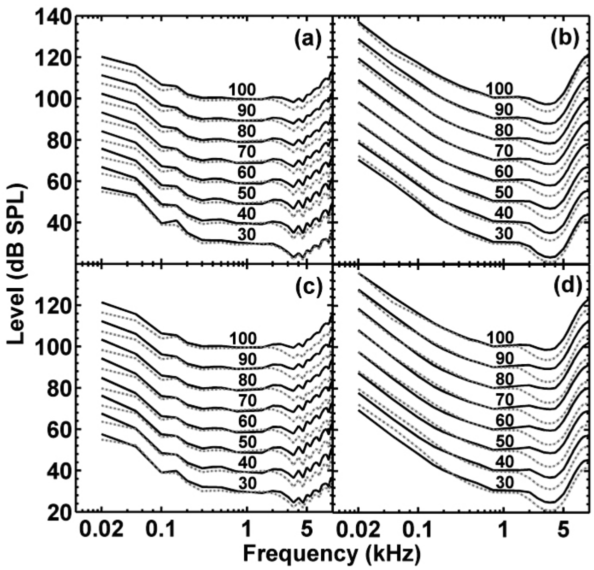

Figure 4. 30 - 100 phon equal-loudness-level contours at 10 dB intervals. Taken from the model of [3] (*a, c*) and [4] (*b, d*), processed with Eq. 2 (*a, b*) and Eq. 3 (*c, d*). Solid line indicates processed signals and dotted line indicates unprocessed signal reference.

The variation in equal-loudness-level contours (Fig. 4) shows similar dependence on frequency and listening-level as in section 1. Maximum effect can be observed in the 50 – 80 dB SPL range, where maximal positive LOE effects at 1 kHz interact with maximal negative LOE effects at other frequencies to produce maximum deviation from the original contours. Maximum deviation of ~6 dB is seen around the 3-kHz range and at very high frequencies (>10 kHz) in Fig. 4(d).

Qualitatively, the nonlinearities have caused a significant flattening of the 'speech range' minima, which is most pronounced in Fig. 4(d). Also, the unprocessed curves of the Glasberg and Moore model appear to conform well to the general features of the equal loudness contours of literature (*e.g.*, [13]), while the derived contours of the Zwicker and Fastl model appear strikingly different to the data of literature.

The recent work of Suzuki and Takeshima [13] analyzed systematic deviation in equal-loudness-level contours throughout the history of the literature. In particular, this study showed large disparity between the data of *early* literature (*e.g.*, Fletcher and Munson [1] and Robinson and Dadson [14]) and that of *recent* literature (*e.g.*, Takeshima et al. [15]). Notably, large deviation in the low frequencies and speech range was shown which is not dissimilar to that produced here. In light of the present study, it might be concluded that the similarity in deviation perhaps implies that there was significant nonlinearity present in the apparatus of the earlier studies.

5. SYNTHESIZED MUSIC NOTES LOE (f, ℓ)

A simple, robust digital additive synthesis model was chosen for the following analysis. The model is based on that described by Horner and Ayers [16] and implemented by Rocamora et al [17]. The model is able to produce musical notes of arbitrary duration for the following instruments: horn, clarinet, oboe, bassoon, flute, piccolo, saxophone, trumpet, tuba and trombone.

250-ms duration musical notes were synthesized at fundamental frequencies between 20 – 5,000 Hz. The signals were then processed using Eq. 3. The loudness model of Glasberg and Moore [4] was used to estimate loudness for the unprocessed and processed signals at simulated listening-levels of 30 – 120 dB SPL. Then the LOE function of frequency was calculated according to the difference in maximum predicted short-term loudness-level of the un-processed and processed signals.

250-ms duration synthesized notes were generated digitally at the sample rate of 44.1 kHz and the over-sampling method described previously was used. The signals were then processed by the over-sampled nonlinear function (Eq. 3). Fig. 5(a) (*upper*) shows the power spectral density for notes generated with an f_0 of 1 kHz for each instrument and Fig. 5(b) (*lower*) shows the effects of the nonlinear processing.

Fig. 6 shows LOE as a function of frequency for synthesized notes in the range 50 – 5,000 Hz and for the various instruments at simulated sound pressure levels of 60 - 100 dB. Strong dependence on frequency is evident for all instruments and wide variation exists among instruments. For all instruments and levels there was a general trend of declining LOE values towards higher frequencies (> 2,000 Hz), which is also evident in the means (dark line), and is consistent with the results of the previous pure-tone experiment. This suggests that repertoire and instrumentation will have a strong influence on the results in the case of music signals.

Frequency-dependent variation in LOE value for any given instrument may be explained in terms of an interaction between spectral content of the signal and variation in masking effects, which results from variation in equivalent rectangular bandwidths.

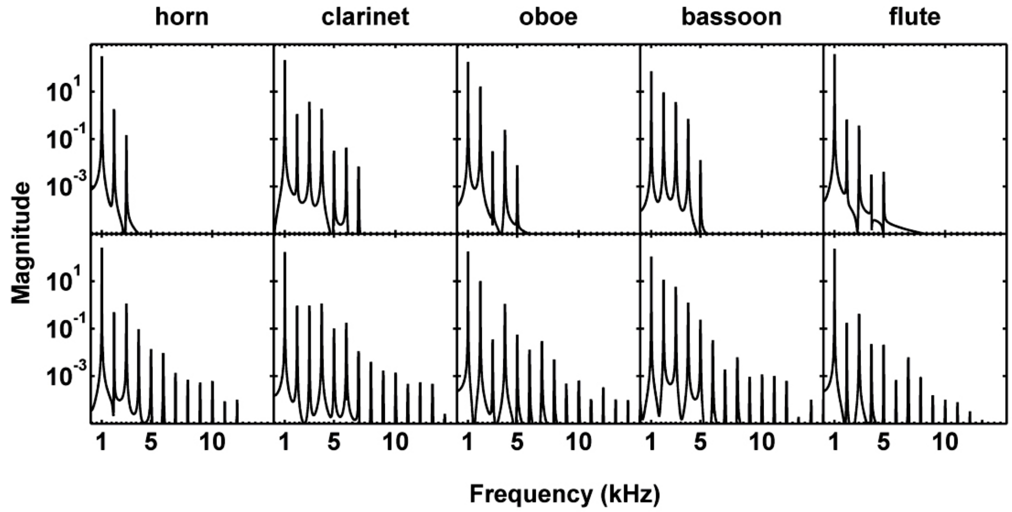

Figure 5. Power spectral density of synthesized musical notes with 1-kHz fundamental frequency. (a) Unprocessed signals (*upper*). (b) Signals processed with Eq. 3 (*lower*).

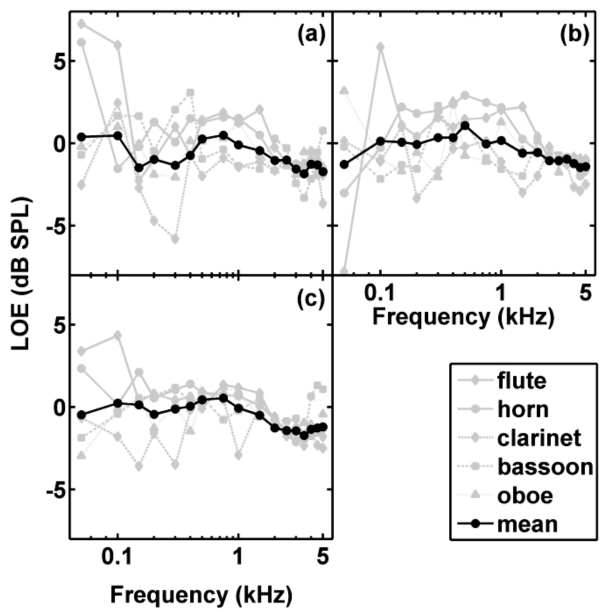

Figure 6. 60, 80 and 100 dB SPL LOE as a function of f_0 for synthesized 250-ms notes, calculated from the results of the model of [4]. (a) 60 dB SPL (*left upper*). (b) 80 dB SPL (*right upper*). (c) 100 dB SPL (*left lower*).

6. EQUAL-LOUDNESS-LEVEL CONTOURS FOR MUSICAL NOTES

The simulated method of adjustment algorithm of experiment 2 was employed using synthesized horn instrument musical note signals of 250-ms. The algorithm recursively adjusted the sound pressure level of the target note by a value inversely proportional to the difference in predicted loudness level of the reference musical note (f_0 1 kHz) until convergence to within 1 dB (phon). The signals were then processed by the nonlinear function (Eq. 3) and the process was repeated.

The variation in equal-loudness-level contours (Fig. 7) shows similar dependence on frequency and listening-level as in experiment 3. Maximum effect can be observed in the 50 – 90 dB SPL range and in the note fundamental frequency range of 3 – 5 kHz, where a maximum deviation of ~5 dB is seen. The qualitative flattening of the speech-range minima is consistent with the results of the previous experiment.

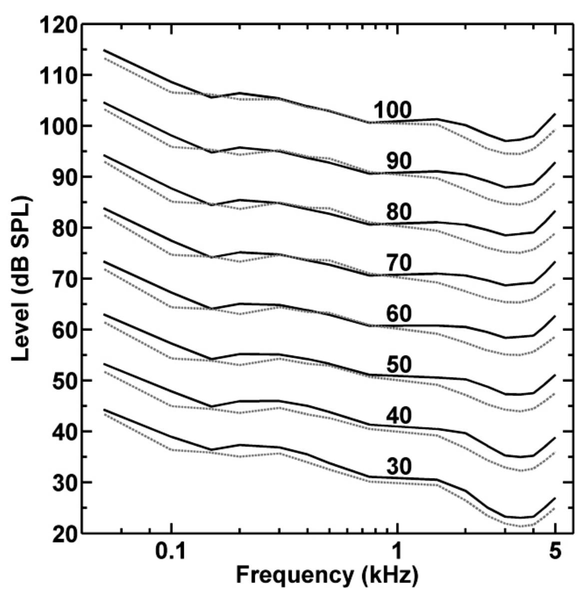

Figure 7. Equal-loudness-level contours for synthesized horn notes at 10 dB intervals for levels between 30 – 100 phon, taken from the model of [4]. Solid line indicates signals processed with Eq. 3 and dotted line indicates unprocessed signal reference.

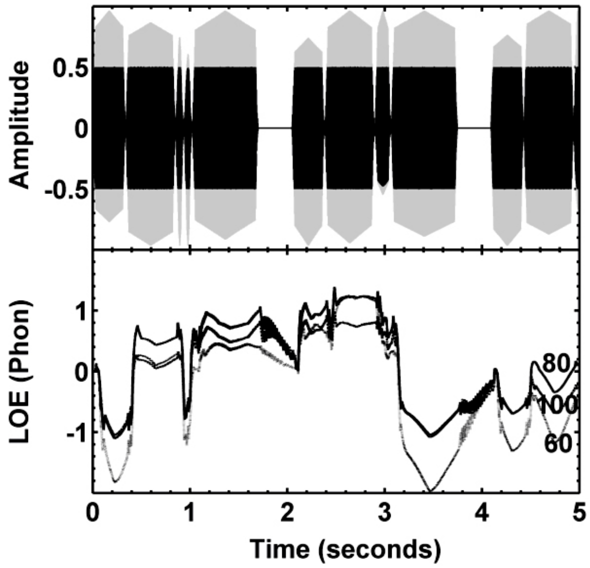

Figure 8. (a) Time-domain representation of the audio signal. Grey describes the unprocessed signal and black describes the signal processed with Eq. 3. (b) 60, 80 and 100 dB SPL LOE as a function of time, calculated from the results of the model of [4].

7. LOE (*ℓ, t*) - TIME-VARYING MONOPHONIC MUSIC

The loudness model of the previous experiment was used to produce LOE as a function of time. Using the synthesis model of the previous experiment, a MIDI file was used to generate a monaural audio signal of an excerpt of monophonic music for the horn instrument. The audio signal was then processed with Eq. 3. The two audio signals were then independently input to the loudness model and a time varying LOE function of listening-level was produced.

The loudness model was input with the synthesized monaural musical signal at simulated peak levels of 60, 80 and 100 dB SPL. The signal was then nonlinearly processed and the difference in short-term loudness was calculated for each unprocessed and processed signal and plotted as a time-varying LOE graph.

The MIDI score comprised a sequence of MIDI notes taken from a musical score. The audio signal was produced digitally at a sample rate of 44.1 kHz and processed using the over-sampled method described previously. The waveform of the nonlinearly processed monaural audio signal is shown in Fig. 18(a), superimposed on the waveform of the original unprocessed signal.

Fig. 8(b) shows the 60, 80 and 100 dB SPL time-varying LOE functions produced, which are correlated in time with the time-domain (waveform) representation of the audio signal in Fig. 8(a). For the highest SPL of 100 dB, the LOE value varies between around -2 and 1.5 dB (phon). This indicates a strong variation in effects depending on the particular note played, which is consistent with the results of section 6. Furthermore, where subsequent musical notes alternate between large negative LOE values and large positive LOE values maximum disruption of the perceived dynamic range can be expected between the notes (*e.g.*, around 3 seconds – Fig. 18(b)).

8. CONCLUSION

A formal framework for LOE has been given. The two arbitrary static nonlinearities have been characterized in terms of LOE for pure-tone and synthesized musical notes, both as a function of frequency and of listening-level and in the equal-loudness-level contours using loudness models [3, 4]. It has been shown that the steeper-sloped nonlinearity of Eq. 3 has produced greater effects on the pure-tone loudness function. In the most extreme case of Eq. 3 at 100 Hz, an approximately 6 dB compression has resulted in a surprising 6 dB *increase* in loudness. Key differences between the models have been discussed and the related limitations of bandwidth with respect to the impact of harmonic distortion products illustrated. The equal-loudness-level contours have illustrated frequency-dependent interactive effects and it has been suggested that systematic differences between the data of early and recent literature might be partially explained by the presence of nonlinearity in the apparatus (*e.g.*, headphones and amplifiers). For the musical notes, the results of the computational analysis show a significant dependence on frequency and instrument. In the case of the horn solo musical excerpt, the time-varying LOE function showed a large effect on the relative loudness difference between notes. Future work should include further extension of the analysis to other types of nonlinearity, other types of musical instruments (*e.g.*, percussion) and should include subjective testing which would provide support for the results of the model and computational analysis. Such measurements might also be useful in the specification and characterization of nonlinear audio systems such as loudspeakers and microphones.

9. REFERENCES

[1] Fletcher, H., and Munson, W.A., "*Loudness, its Definition, Measurement and Calculation*", J. Acoust. Soc. Am. 5, 82–108, 1933.

[2] Simpson, A.J.R., "*Auditory Perception of Dynamic Range in the Nonlinear System*", AES Convention: 128, Paper number: 8103, May 2010.

[3] Zwicker, E., Fastl, H., "*Psychoacoustics: Facts and Models*", 2nd Edition, Springer-Verlag, Berlin, 1999, Chap. 8.

[4] Glasberg, B., Moore, B., "*A Model of Loudness Applicable to Time-Varying Sounds*", J. Audio Eng. Soc., Vol. 50, No. 5, pp. 331–342. May 2002.

[5] Klippel, Wolfgang, "*Tutorial: Loudspeaker Nonlinearities--Causes, Parameters, Symptoms*", JAES Volume 54 Issue 10 pp. 907-939; October 2006.

[6] Hibbing, Manfred; Griese, Hans-Joachim, "*New Investigations on Linearity Problems of Capacitive Transducers*", AES Convention: 68, Paper number: 1752, March 1981.

[7] Moore, B.C.J., Glasberg, B.R., Stone, M.A., "'*Why Are Commercials so Loud?' Perception and Modeling of the Loudness of Amplitude-Compressed Speech*", JAES Volume 51 Issue 12 pp. 1123-1132; December 2003.

[8] Schetzen, M., "*The Volterra and Wiener Theories of Nonlinear Systems*", Robert Krieger Publishing, 1980.

[9] Zwicker, E., and Fastl, H., "*Psychoacoustics, Facts and Models*", Berlin: Springer Verlag, 1990.

[10] Glasberg, B.R., and Moore, B.C.J., "*Derivation of Auditory Filter Shapes from Notched-Noise Data*" Hear. Res., vol. 47, pp. 103–138, 1990.

[11] Smith, J.O. III., and Abel, J.S., "*Bark and ERB bilinear transforms*", Speech and Audio Processing, IEEE Transactions on, vol.7, no.6, pp.697-708, Nov 1999.

[12] Kingsbury, B.A., ''*A direct comparison of the loudness of pure tones,"*, Phys. Rev. 29, pp. 588–600. 1927.

[13] Suzuki, Y., and Takeshima, H., "*Equal-Loudness-Level Contours for Pure Tones*", J. Acoust. Soc. Am., vol. 116 (2), pp 918–933, 2004.

[14] Robinson, D. W., and Dadson, R. S. "*A re-determination of the equal-loudness relations for pure tones*", Br. J. Appl. Phys. 7, 166–181, 1956.

[15] Takeshima, H., Suzuki, Y., Ashihara, K., and Fujimori, T., ''*Equal-Loudness Contours Between 1 kHz and 12.5 kHz for 60 and 80 Phons*'', Acoust. Sci. Tech. 23, 106–109, 2002.

[16] Horner, A., and Ayers, L., "*Modeling Acoustic Wind Instruments with Contiguous Group Synthesis*", J. Audio Eng. Soc., 46(10):868–879, 1998.

[17] Rocamora, M., Lopez, E., Jure, L., "*Wind Instruments Synthesis Toolbox for Generation of Music Audio Signals with Labeled Partials*", SBCM 2009 – 12th Brazilian Symposium on Computer Music, Brasil, Sept. 2009.