ARCHIVES OF ACOUSTICS
36, 2, 239–250 (2011)
DOI: 10.2478/v10168-011-0019-7

Loudness Assessment of Musical Tones Equalized in *A*-weighted Level

Despina KLONARI[(1)], Konstantinos PASTIADIS[(1)], Georgios PAPADELIS[(1)], George PAPANIKOLAOU[(2)]

Aristotle University of Thessaloniki
Faculty of Fine Arts
[(1)] *Department of Music Studies*
[(2)] *Department of Electrical and Computer Engineering*
541 24 Thessaloniki, Greece
e-mail: {klonari, pastiadi, papadeli}@mus.auth.gr
pap@eng.auth.gr

(*received January 28, 2011; accepted April 8, 2011*)

The present study was carried out to determine whether recorded musical tones played at various pitches on a clarinet, a flute, an oboe, and a trumpet are perceived as being equal in loudness when presented to listeners at the same *A*-weighted level. This psychophysical investigation showed systematic effects of both instrument type and pitch that could be related to spectral properties of the sounds under consideration. Level adjustments that were needed to equalize loudness well exceeded typical values of JNDs for signal level, thus confirming the insufficiency of *A*-weighting as a loudness predictor for musical sounds. Consequently, the use of elaborate computational prediction is stressed, in view of the necessity for thorough investigation of factors affecting the perception of loudness of musical sounds.

Keywords: loudness perception, *A*-weighted level equalization, wind instrument sounds.

1. Introduction

Loudness is one of the fundamental attributes of auditory sensation. The problem of loudness estimation and control of speech and musical sounds has long been an issue in various applications of audio engineering and technology. A common and standardized method that could accurately calculate loudness for a variety of sounds is also desirable in contemporary psychoacoustic and sound quality research, especially in cases where loudness equalization of experimental stimuli is necessary so that any loudness-related confounding effects in the measurement of various perceptual attributes of sound other than loudness (e.g. pitch, timbre) have to be eliminated (Kostek, Wieczorkowska, 1997; Kostek, Czyżewski, 2001; Skovenborg, Nielsen, 2004; Papanikolaou, Pastiadis, 2009). This necessity has led researchers to adopt frequently used objective measures, such as A-weighted level and other measures derived from spectral loudness summation models. Traditionally, the audio engineering community has used the A-weighted sound pressure level as an indicator of loudness, particularly for noise signals. However, in several cases, its use has been strongly criticized (see for example Hellman, Zwicker, 1987).

In the case of musical sounds, gradations in playing level (i.e. from pianissimo to fortissimo), tonal variations within a pitch range (i.e. from low to high) and variations in spectral envelope influence the perceived loudness in various ways. This influence also differs considerably among various types of musical instruments (Miśkiewicz, Rakowski, 1994). Measures of loudness derived from controlled listening experiments may differ substantially from those obtained by objective methods. However, loudness equalization of musical tones by means of listening experiments may suffer from a number of problems. First, the subjective nature of human perception manifests itself by relatively large individual differences in equal loudness estimations obtained from different listeners, even when they are well-trained (Hajda *et al.*, 1997). Second, the responses of listeners are susceptible to judgmental bias related with the type of measurement (Silva, Florentine, 2006). As a consequence, recent psychoacoustic research has elaborated strategies for the control of loudness of musical sounds, based on combinations of both subjective and objective techniques. These combinations include A-weighting or estimates provided by computational loudness models, together with data drawn from exploratory listening experiments. Examples outlining the range and the variety of different approaches to loudness equalization of musical sounds can be drawn mainly from timbre-related research. Several studies in this area (Grey, 1975; Kendall, Carterette, 1991; 1993; Iverson, Krumhansl, 1993; Iverson, 1995; Sandell, 1995; Kendall *et al.*, 1999; Marozeau *et al.*, 2003) have utilized various schemas of preliminary A-weighting and/or subsequent subjective equalization. However, in the majority of these studies equalization of loudness was based on a rather small number of listeners.

The aim of this study is to investigate and quantify to a larger extent the possible confounding effects of the type of instrument and pitch on the judgments of loudness equivalence of musical tones played on selected instruments of the wind family. Since A-weighting is frequently used in psychoacoustical research, our investigation also utilizes it as a measure against which loudness judgments are compared.

Four instruments from the wind section of the symphony orchestra were chosen: Bb clarinet, flute, oboe and Bb trumpet. These instruments are used frequently in psychoacoustic experiments and play a major role in orchestration. Even though they belong to the same family of musical instruments, they considerably differ in timbre. Despite these differences, they all show a well-defined steady-state portion which makes them appropriate for our study. We also targeted on four pitch values from the upper part of the medium register of musical pitches – A4 (440 Hz), C#5 (554 Hz), A5 (880 Hz) and C#6 (1109 Hz) – which define intervals frequently used in orchestration of Western culture music, such as the unison, the major third, the octave and the compound major third (octave plus a major third). As loudness judgments may be susceptible to bias effects due to individual differences and factors related to experience, motivation, training and attention of the listener (Moore, 1989) our work employed a considerably larger number of participants, in comparison to the majority of previous studies which used listening procedures for loudness equalization of musical sounds. The pool of subjective data collected through this study were finally compared to A-weighting equalization of loudness and any discrepancies are explored in detail.

2. Method

2.1. Design

Single notes from each instrument at A4, C#5, A5, C#6 – one note per pitch value – were used as test stimuli. Listeners compared the loudness of different levels of each of the 16 test tones (4 instruments × 4 pitch values) to that of an A4 tone, played on an oboe at a level of 70 dBA, which was selected as the standard tone. Loudness equivalence measures were finally expressed in terms of obtained L_{Aeq} values for each combination of instrument and pitch. Each of these values indicates the level of the corresponding test tone that led to an equal loudness judgment when compared to that of the standard.

2.2. Stimuli and apparatus

Single notes were played by experienced music performers on each instrument used in this study, at all four targeted pitch values. The duration of each note was 3.5 sec. The recordings were made at the Laboratory of Electroacoustics and Television Systems (Aristotle University of Thessaloniki) with the use of an AKG C460B-CK61 microphone and a Pro Tools/HD2 system. The microphone was placed in front of the performer at a height of 155 cm, at an average distance of 90 cm from the instrument. The instruments were initially tuned with the use of an electronic tuner, then each one was separately aligned with the oboe's tuning. A final tuning alignment was carried out having all instruments play together as an ensemble.

A major concern during the recording session was to ensure that the recorded tones sounded as natural as possible within the desired range of sound levels. For that reason, the level of each tone was monitored through a Brüel & Kjaer 2230 sound level meter at a distance of 1 m in front of the instrument, and musicians were asked to adjust their playing level so that the tone's A-weighted level was as close as possible to the 70-dB standard level. Subsequently, the tones were played back through the earphones used in the listening sessions and the tone levels were fine adjusted to the standard with the use of a Brüel & Kjaer 4128 Head and Torso Simulator (HATS).

2.3. *Participants*

Forty-eight musicians, 22 males/26 females, aged 19–44, mostly students from the School of Music Studies (Aristotle University of Thessaloniki) participated voluntarily in the experiment. Six of them had previous experience with psychoacoustic testing procedures. None of the participants reported any diagnosed hearing disorder.

2.4. *Loudness-matching procedure*

Loudness judgments were performed by listening to pairs of the recorded tones. A basic set of 16 paired stimuli was created from all possible combinations of instruments and pitches as follows: Each pair consisted of an oboe A4 tone (standard stimulus) followed by one of the 16 test stimuli. The pair "Ob/A4 – Ob/A4" was also included in the set. Stimuli within each pair were separated by 1 sec of silence. As the order of stimuli presentation may profoundly affect loudness judgments (Silva, Florentine, 2006), the same paired sounds were also presented in reverse order within each individual pair. Thus, the total number of paired stimuli was extended to 32, each containing the standard stimulus (oboe A4 @ 70 dBA), either at the first or at the second position within a presentation pair.

Loudness matches within each pair were obtained through an adaptive two alternative forced-choice procedure based on a Parameter Estimation by Sequential Testing (PEST) paradigm (Taylor, Creelman, 1967). PEST uses a combination of increasing and decreasing stimulus steps (levels) along a block of trials that change both in direction and step size (tracking algorithm) according to the participant's responses and predefined rules for ending the measurement (Leek, 2001).

In our case, PEST was realized similarly as described in literature (Gelfand, 2004). Level variations were allowed on only one tone – called the variable tone – within each paired stimuli, either the first or the second. The level of the non-varied tone – called the fixed tone – was set at 70 dB L_{Aeq}. An initial level difference of ± 8 dB was used as being sufficient to elicit a distinct loudness mismatch between the variable and the fixed tone, thus leading to either a downward or an upward adjustment of the variable tone. Positive vs. negative initial level differences of the variable tone as well as the order of presentation of the fixed and variable tones were balanced across tracks (Fig. 1). The initial step size for

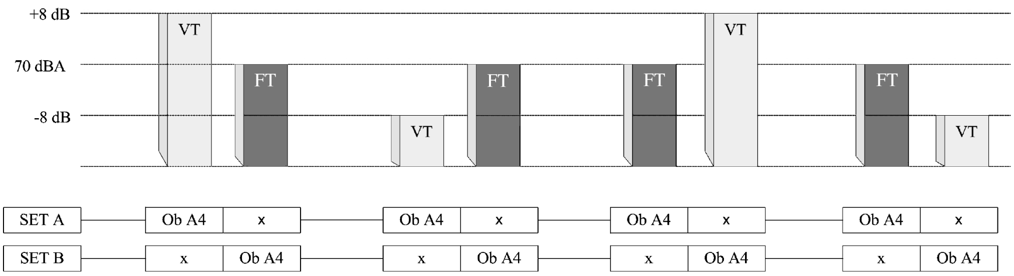


Fig. 1. Combinations of within-pair sequential order and initial level difference for a PEST track that were used for presenting each of the 16 paired stimuli drawn from the basic set (VT: Variable Tone, FT: Fixed Tone, x: One of the 16 test stimuli, SET A: pair in direct sequential order, SET B: pair in reverse sequential order).

level variation was selected at 2 dB. The step size was doubled after two consecutive identical responses, while a change of response caused both a reversal of the direction of testing level variation and halving of the step size. The convergence threshold was set to 1 dB. The track terminated when the step size was below the convergence threshold and the level for equal loudness was calculated as the testing level of a next trial. An additional termination rule was also imposed such that the number of trials on a track did not exceed 15. The procedure converged at the 50% point of the psychometric function.

During the main listening sessions participants were seated in a sound isolated booth close to the room where the experimenter sat in, and were presented with 128 pairs of stimuli in random order. Sound stimuli were presented diotically through a pair of circumaural earphones (Sennheiser HD 545). After each presentation of a pair the listener responded "YES" when the second tone within the pair was louder than the first one, and "NO" in any other case (i.e. if the second tone was less or equally loud). In most tracks, the procedure converged in less than seven responses for each pair, and the whole session was usually completed in less than 2 h 15 min (necessary breaks were also included). Proper instructions were given by the researchers and a simple test of acquaintance with the experimental procedure preceded the main psychoacoustical testing.

All tests were conducted with the use of specialized software – developed in National Instruments' LabVIEW suite – employing stimuli randomization, hardware control, signal presentation, realization of the PEST procedure, recording of subjects' responses, session timing control, and data storage/administration.

3. Results and discussion

Distributions of equal loudness judgments for each of the assessed combinations of instrument and pitch are shown in Fig. 2 in terms of the corresponding sound level difference between the level of each test tone and that of the standard tone (oboe A4 @ 70 dBA re 20 μPa). Figure 2 also shows the respective means

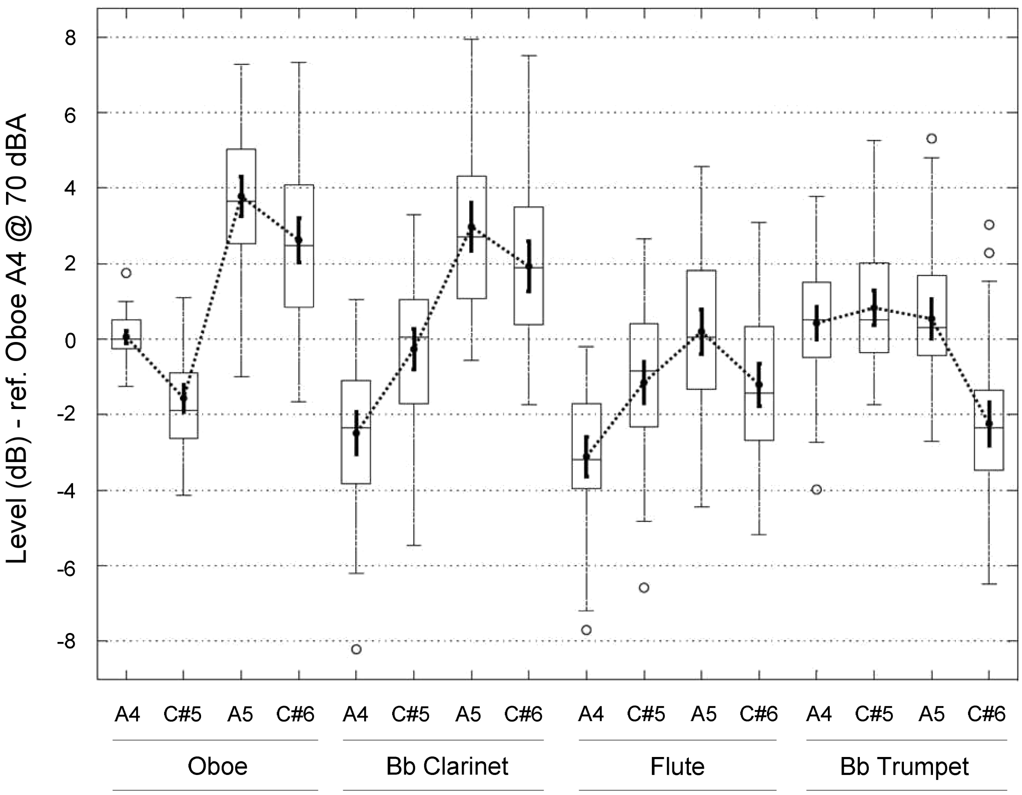


Fig. 2. Distribution of level differences between each experimental condition (combination of instrument type and pitch) and that of the standard tone (oboe A4 @ 70 dBA re 20 μPa) for the equally-loud condition. Box-and-whisker plots show distributions within each subgroup of data. Error bars within each box show mean level differences with 95% confidence intervals. Dashed lines indicate patterns of variation across pitch range within each instrument type.

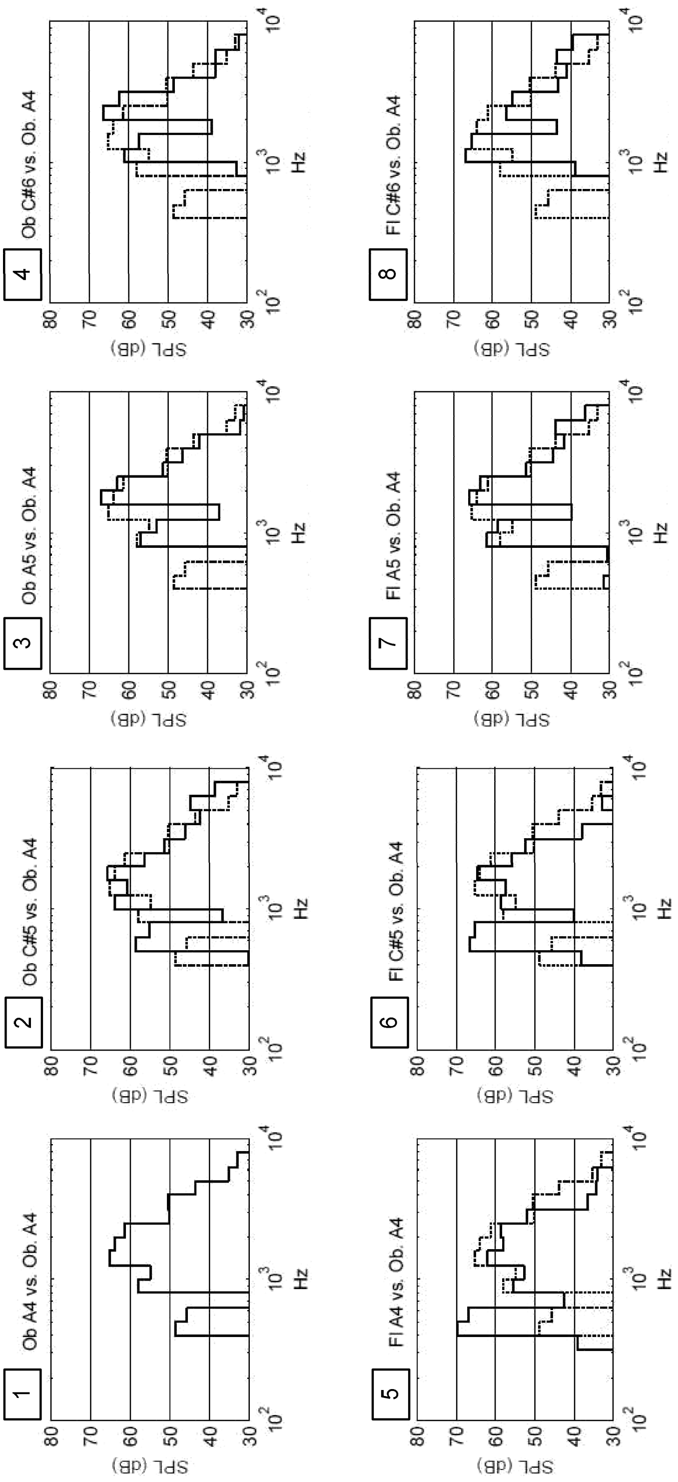
and 95% confidence interval for each instrument-pitch combination. Interestingly, when the test tone was an oboe A4, the same as the standard, the median and mean values were nearly 0 dB and the range of equal loudness judgments reached its minimum value (3 dB), indicating the relatively high reliability of the measurement procedure. On the contrary, noticeable contrasts were found between the values of median/range obtained for the "oboe A4 – oboe A4" pair and those for the majority of the other pairs, which illustrate an apparent influence of pitch and instrument on loudness judgments. Even when the type of instrument or pitch was the same for the test and the standard stimuli, equal loudness judgments displayed pronounced differences from those obtained for the "oboe A4 – oboe A4" pair. For example, a relatively large level difference (mean: 3.8 dB) and considerable variability in equal loudness judgments among participants (range: 7.3 dB), was obtained for the "oboe A4 – oboe A5" pair, despite the fact that both tones were produced by the same instrument.

The main effect of instrument and pitch, as well as their interaction on equal loudness judgments was examined using a 4 (instrument type) × 4 (pitch) repeated-measures ANOVA (Kiess, Bloomquist, 1985). The instrument had a significant influence on equal loudness judgments $F(2.56, 120.75) = 68.85$, $p < 0.001$, with a large effect size for clarinet tones, $F(1, 47) = 18.87$, $r = 0.53$, flute, $F(1, 47) = 162.23$, $r = 0.88$, or trumpet $F(1, 47) = 91.47$, $r = 0.81$. There also was a significant main effect of pitch on equal loudness judgments $F(1.93, 90.78) = 121.52$, $p < 0.001$, with important contrasts between the major third $F(1, 47) = 28.83$, $r = 0.61$, octave $F(1, 47) = 214.26$, $r = 0.9$, and a compound major third $F(1, 47) = 50.53$, $r = 0.71$ higher to the test tone, compared to the unison (A4) presentations. A significant interaction effect between the type of instrument and pitch $F(6.82, 320.57) = 122.12$, $p < 0.001$ was also observed.

In several cases, the mean level differences between the test tones and the standard oboe tone were considerably larger than 0.5–1 dB (see Fig. 2), which is reported as the size of the just noticeable difference (jnd) for level, at least for pure tones and noises (Moore, 1989; Zwicker, Fastl, 1990; Campbell, Greated, 2001). Therefore, those differences should be regarded as perceptually important and provide further evidence on the inappropriateness of A-weighting as an index of perceived loudness, especially at levels higher than those for which this index was originally specified (Hellman, Zwicker, 1987).

The patterns of level differences seen in Fig. 2 could partially be explained by the acoustical characteristics of the tones especially those associated with perceptually important aspects of the spectral energy distribution of the sound, such as the strength of the fundamental, the total number of overtones (and thus the total bandwidth, especially at very high pitches), the existence of major spectral energy concentrations (formants), or their position within an instrument's playing range (register) at which these notes are located. In order to explore in detail the effect of the spectral envelope of each individual experimental tone on the judgment of its loudness, 1/3-octave-band SPLs were plotted for each tone separately against that of the standard (Fig. 3). A comparison of the two spectral envelopes on each graph in Fig. 3 demonstrates that, despite the fact that both sounds have the same overall A-weighted SPL, differences between 1/3-octave spectral envelopes may result in differences in loudness, similar to those shown in Fig. 2 (see also Zwicker, Fastl, 1990).

For example, at A4, a strong character of the fundamental in the flute and in the clarinet, as contrasted to the oboe, (Figs. 3.5 and 3.9) could be the cause of the corresponding mean level differences seen in Fig. 2. In contrast, spectral similarities between the trumpet and the oboe (Fig. 3.13), namely the existence of a weaker fundamental against higher partials, due to the so-called spectrum transformation function (Benade, 1991), and main formant prominence at around 1.2–1.5 kHz in both instruments (Meyer, 2009), associate with the lowest difference in mean level in the equal loudness condition. A consistent loudness reduction with increasing pitch for the clarinet, the flute, and the oboe – with only



[Fig. 3. Parts 1–8.]

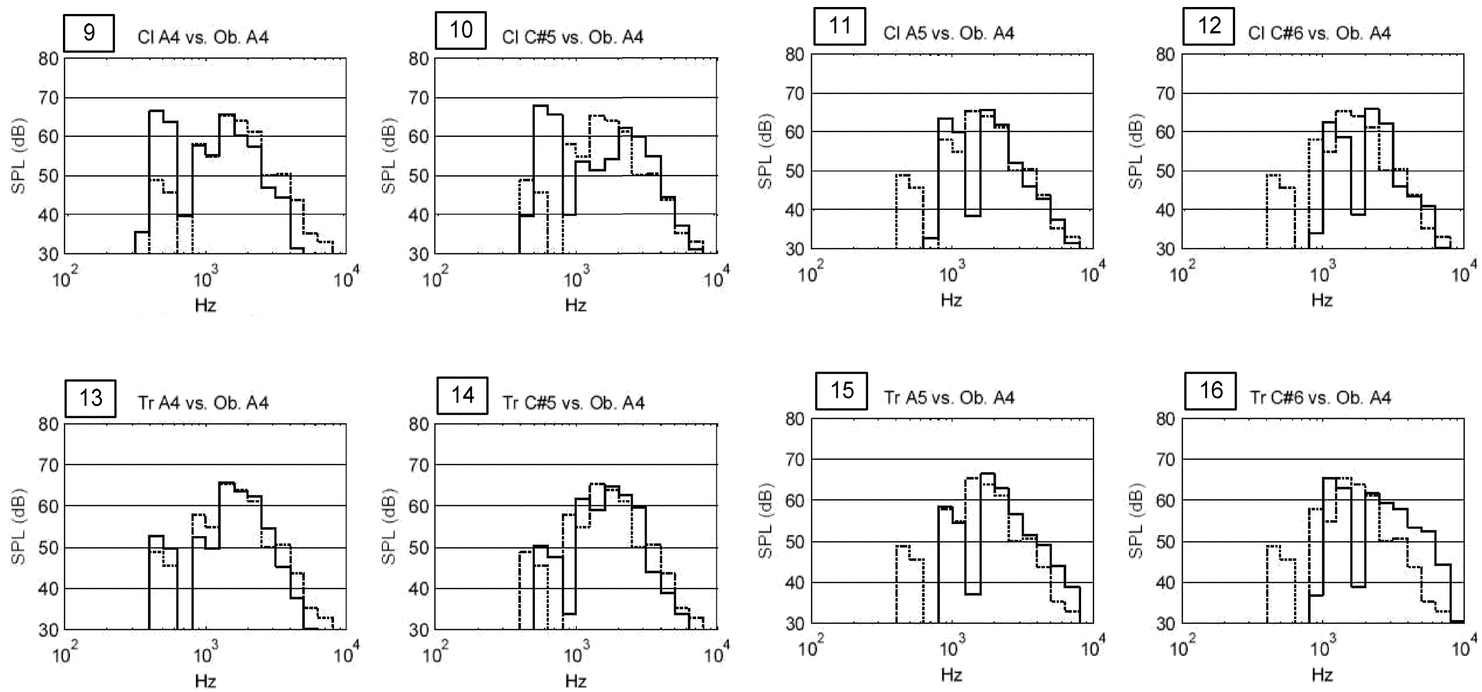


Fig. 3. 1/3-octave-band SPLs of the tones used in this work. Each graph shows the spectrum of a single tone (solid line) and the spectrum of the standard tone (oboe A4 @ 70 dBA re 20 µPa – dotted line). All tones were set to an overall sound level of 70 dBA.

a slight decrease in the higher register – can be observed. As all three instruments show well-developed overtones, elimination of the 440 Hz-band contributing to loudness with increasing pitch could potentially account for the observed level increase required to equalize the test tone and the standard in loudness. In cases where overtones in these instruments become increasingly rich, as for example in the trumpet, C#6 where they extend up to 8–9 kHz (Fig. 3.16), these overtones counterbalance the energy lack at the 440 Hz region and lead to a noticeable level reduction displayed in the trumpet's pattern toward C#6 (Fig. 2). In the case of the C#5 tone of the oboe deviation from the pattern of monotonic level increase towards higher pitch values could be attributed to an enhancement of the fundamental by a sub-formant located in a region of ca. 550–600 Hz (Fig. 3.2). For the trumpet, a combined effect of the higher overtone prominence and an enhancement of the formant structure of the instrument, may account for the minimal positive level differences in the cases of A4, C#5, and A5 tones. The above explanatory approaches on the observed variations of loudness and their association to the degree to which the spectral envelope of the musical tones varies with pitch along different musical instruments are consistent with previous findings of Miśkiewicz and Rakowski (1994), who showed that predicted loudness levels of musical instrument tones "are markedly influenced by the variations in spectral envelope that arise from gradations in playing level".

In summary, the present findings lead to the conclusion that, at least in wind-instrument sounds, a frequency weighting other than the typically used *A*-weighting may lead to level differences eliciting loudness equivalence that could differ substantially from the ones observed in this study. To determine the relevance of various types of spectral weighting for prediction of the loudness of musical tones and gain more thorough understanding of the effect on loudness of other acoustical factors, such as temporal envelope of sounds, etc. (Harajda *et al.*, 1993; Meyer, 2009), investigations should be extended over broad ranges of musical dynamic, gradations pitches and instrument timbres. From a methodological point of view, our results point out the importance of integrating both psychoacoustical testing and computational procedures in loudness estimation (Miśkiewicz, Rakowski, 1992; Skovenborg, Nielsen, 2004).

4. Conclusions

The current study concentrated on the problem of loudness equivalence between tones produced by wind musical instruments, presented to the listeners at the same *A*-weighted level. Results demonstrated the significance of both independent and combined effects of the type of instrument and pitch on the judgments of loudness equality. These findings may be of major importance, especially for psychoacoustical experiments that employ musical sounds, where elimination of confounding effects of loudness is necessary. At least in cases of wind instrument tones that lie within intervals of up to a compound major third (octave plus a major third) in the upper part of the medium register of musical pitches – such as those of the current study – alignment of loudness based on A-weighting SPLs may require further adjustments over a range of more than 5 dB. It is therefore evident that physical attributes of the tone, such as its spectral content, together with emerging perceptual aspects, such as pitch and timbre, are the cause of the discrepancy between equal A-levels and levels of equal loudness, and this discrepancy appears to be consistent and predictable. However, it is not easy to come to definite conclusions about the relative importance of various types of factors (i.e. acoustic, perceptual, and cognitive) that contribute to the observed results, since these groups of factors coexist and interact in a complex manner.

A comparison of the present results against those provided by computational models of loudness estimation would provide valuable information about the degree to which widely used models may be valid for loudness equalization of musical sounds.

Acknowledgments

The authors are grateful to the instrument players that participated with enthusiasm in the sound recording sessions. Special thanks are also given to the students and staff of the Department of Music Studies for their voluntary participation to this study. Finally, we would like to take this opportunity to thank the two anonymous reviewers who have donated their expertise and time to reading and commenting on a previous version of this article.

References

1. Benade A.H. (1991), *Fundamentals of musical acoustics*, New edition, Dover Publications Inc., New York.
2. Campbell M., Greated C. (2001), *The musician's guide to acoustics*, OUP Oxford, New York.
3. Gelfand S. (2004), *Hearing: an introduction to psychological and physiological acoustics, fourth edition, revised and expanded*, 5th edition, Informa HealthCare, UK.
4. Grey J. (1975), *An exploration of musical timbre*, Ph.D. Thesis, Stanford University, Stanford.
5. Hajda J., Kendall R., Carterette E., Harschberger M. (1997), *Methodological issues in timbre research*, [in:] *Perception and cognition of music*, Deliège I., Sloboda J. [Eds.], pp. 253–306, Psychology Press, East- Sussex, UK.
6. Harajda H., Mikiel W., Fedyniuk P. (1993), *Microstructure of sound: formants in the dynamical spectra of violin sounds*, Archives of Acoustics, **18**, 1, 17–32.
7. Hellman R., Zwicker E. (1987), *Why can a decrease in db(a) produce an increase in loudness?*, The Journal of the Acoustical Society of America, **82**, 5, 1700–1705.
8. Iverson P. (1995), *Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes*, Journal of Experimental Psychology, Human Perception and Performance, **21**, 4, 751–63.
9. Iverson P., Krumhansl C.L. (1993), *Isolating the dynamic attributes of musical timbre*, The Journal of the Acoustical Society of America, **94**, 5, 2595–603.

10. Kendall R.A., Carterette E.C. (1991), *Perceptual scaling of simultaneous wind instrument timbres*, Music Perception, **8**, 4, 369–404.

11. Kendall R., Carterette E. (1993), *Identification and blend of timbres as a basis for orchestration*, Contemporary Music Review, **9**, 51–67.

12. Kendall R.A., Carterette E.C., Hajda J.M. (1999), *Perceptual and acoustical features of natural and synthetic orchestral instrument tones*, Music Perception, **16**, 3, 327–363.

13. Kiess H., Bloomquist D. (1985), *Psychological Research Methods: A Conceptual Approach*, Allyn and Bacon, Inc., Newton, Massachusetts.

14. Kostek B., Czyżewski A. (2001), *Representing musical instrument sounds for their automatic classification*, Journal of the Audio Engineering Society, **49**, 9, 768–785.

15. Kostek B., Wieczorkowska A. (1997), *Parametric representation of musical sounds*, Archives of Acoustics, **22**, 1, 3–26.

16. Leek M.R. (2001), *Adaptive procedures in psychophysical research*, Perception & Psychophysics, **63**, 8, 1279–1292.

17. Marozeau J., De Cheveigné A., McAdams S., Winsberg S. (2003), *The dependency of timbre on fundamental frequency*, The Journal of the Acoustical Society of America, **114**, 5, 2946–2957.

18. Meyer J. (2009), *Acoustics and the performance of music: manual for acousticians, audio engineers, musicians, builders of musical instruments and architects*, 5th edition, Springer, New York.

19. Miśkiewicz A., Rakowski A. (1992), *Music vs noise: a comparison of loudness estimates*, Archives of Acoustics, **17**, 2, 203–216.

20. Miśkiewicz A., Rakowski A. (1994), *Loudness level versus sound-pressure level: a comparison of musical instruments*, The Journal of the Acoustical Society of America, **96**, 6, 3375–3379.

21. Moore B.C.J. (1989), *An introduction to the psychology of hearing*, 3rd edition, Academic Press, London.

22. Papanikolaou G., Pastiadis C. (2009), *Multiple dichotomies in timbre research*, Archives of Acoustics, **34**, 2, 127–145.

23. Sandell G.J. (1995), *Roles for spectral centroid and other factors in determining "blended" instruments pairings in orchestration*, Music Perception, **13**, 2, 209–246.

24. Silva I., Florentine M. (2006), *Effect of adaptive psychophysical procedure on loudness matches*, The Journal of the Acoustical Society of America, **120**, 4, 2124–2131.

25. Skovenborg E., Nielsen S. (2004), *Evaluation of different loudness models with music and speech material*, Proceedings of 117th Audio Engineering Society Convention, San Francisco.

26. Taylor M.M., Creelman C.D. (1967), *Pest: efficient estimates on probability functions*, The Journal of the Acoustical Society of America, **41**, 4A, 782–787.

27. Zwicker E., Fastl H. (1990), *Psychoacoustics: facts and models*, 2nd edition, Springer, Berlin/Heidelberg.