Preparation of stimuli for timbre perception studies

Ilse B. Labuschagne and Johan J. Hanekom
Bioengineering Research Group, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Hatfield, 0020, South Africa

(Received 2 April 2012; revised 28 May 2013; accepted 19 July 2013)

Stimuli used in timbre perception studies must be controlled carefully in order to yield meaningful results. During psychoacoustic testing of individual timbre properties, (1) it must be ensured that timbre properties do not co-vary, as timbre properties are often not independent from one another, and (2) the potential influence of loudness, pitch, and perceived duration must be eliminated. A mathematical additive synthesis method is proposed which allows complete control over two spectral parameters, the spectral centroid (corresponding to brightness) and irregularity, and two temporal parameters, log rise-time (LRT) and a parameter characterizing the sustain/decay segment, while controlling for covariation in the spectral centroid and irregularity. Thirteen musical instrument sounds were synthesized. Perceptual data from six listeners indicate that variation in the four timbre properties mainly influences loudness and that perceived duration and pitch are not influenced significantly for the stimuli of longer duration (2 s) used here. Trends across instruments were found to be similar.
[http://dx.doi.org/10.1121/1.4817877]

PACS number(s): 43.75.Zz, 43.66.Jh [MAH] Pages: 2256–2267

I. INTRODUCTION

Music perception studies often focus on rhythm perception, pitch perception (including melodic contours and harmony), and timbre (Jackendoff and Lerdahl, 2006; Krumhansl, 2000; Limb, 2006; Peretz and Coltheart, 2003; Peretz and Hyde, 2003). Timbre is generally defined as a complex and multidimensional quality of sound that allows a listener to discriminate between two sounds of equal loudness and pitch (ANSI, 1994). While timbre perception is important in music perception, e.g., timbre discrimination and recognition assists in the identification of emotional intent and structural organization of a musical work (Gfeller *et al.*, 1998), timbre discrimination also assists in the segregation of auditory streams (Cusack and Roberts, 2004; Galvin *et al.*, 2007; Galvin *et al.*, 2009).

Since timbre is a multidimensional property of sound (Caclin *et al.*, 2005; Grey, 1977; Grey and Gordon, 1978; Kong *et al.*, 2011; McAdams *et al.*, 1999; McAdams, *et al.*, 1995), an investigation of the constituents of timbre may be useful. A systematic investigation of timbre constituents may, for example, be used to identify specific problem areas for groups that experience difficulty in timbre perception (Limb, 2006).

Original instrument recordings or modified recordings of instruments have been used previously to study timbre perception. These stimuli are useful to represent real-world situations. Examples include Gfeller *et al.* (1998) who used original recordings in instrument identification tasks to compare the performance of normal-hearing (NH) listeners and cochlear implant users; Emiroglu (2007) and Emiroglu and Kollmeier (2008) who investigated just noticeable differences (jnds) of instruments that were similar in two out of three timbre properties; studies that used recordings in which individual spectral harmonics have been modified to determine spectral distortion discrimination (Gabrielsson and Sjögren, 1971; Gunawan and Sen, 2008).

While original recordings are valuable in timbre perception studies, the use of synthesized sounds recreated to incorporate specific timbre properties will allow a systematic evaluation of timbre perception. Familiar methods for instrument synthesis include physical modeling, wavetable synthesis, and additive, subtractive, and multiplicative synthesis (De Poli, 1983; Fletcher and Rossing, 1999). Physical modeling represents a real-world instrument as a set of differential equations describing the physical properties of an instrument, making high fidelity sound reproduction possible. However, direct manipulation of timbre properties may be difficult or impossible as physical model parameters generally do not relate simply to the timbre properties. Wavetable synthesis uses a characteristic wave segment of an instrument recording that can be replayed continuously to produce any note duration. Direct timbre manipulation using wavetable synthesis requires individual adjustments of spectral and temporal properties of each recording segment, which is unfeasible in an adaptive procedure that may be used in a perception study. Real-time adjustments of timbre properties for use in adaptive procedures and explicit control of temporal and spectral timbre properties is only practically realizable using additive synthesis.

Simple additive synthesis models are typically used in perception studies and usually include the following three properties, or equivalents thereof: (1) brightness, a spectral property associated with the spectral centroid of a sound, (2) irregularity, a spectral property concerning the difference in magnitude of subsequent harmonic partials, and (3) the temporal envelope, which usually consists of descriptions of attack, sustain, decay, and end segments (Caclin *et al.*, 2005; Jensen, 1999b; Krimphoff *et al.*, 1994). More sophisticated models contain additional timbre properties, like shimmer

 0001-4966/2013/134(3)/2256/12/$30.00

(variations in the temporal envelope magnitude and called tremolo in musical terms), jitter (variations in the frequency of partials and called vibrato in musical terms), stretched harmonics (where each subsequent harmonic may not be a perfect integer multiple of the fundamental frequency, F_0), spectral flux (changes in spectral envelope from the start to end times of a sound), and noise (for example, the noise contributing to the "breathy" nature of a flute) (Jensen, 1999b, 2001; McAdams *et al.*, 1999). These additional timbre properties are useful to achieve more life-like sounds during synthesis.

During the development of an additive synthesis model incorporating a selected set of timbre properties care must be taken to monitor the set as a whole when varying a particular timbre parameter, since a change in one parameter may produce changes in other parameters. For example, Caclin *et al.* (2005) used synthetic tones in dissimilarity rating experiments where the LRT, spectral centroid, spectral flux, and attenuation of even harmonics were varied. However, since all of these timbre properties are not independent, a spectral centroid shift was observed when investigating the perception of attenuation of even harmonics, potentially influencing perception data. This observation underscores that an appropriate synthesis model should be defined in a way that allows a specific timbre parameter to be varied independently from others.

Apart from the possibility of one timbre parameter influencing another, loudness, pitch, and duration cues must also be eliminated before experimentation (Grey, 1975). Although most studies balance sounds for pitch, loudness, and duration (Grey, 1977; Grey and Gordon, 1978; McAdams *et al.*, 1995), some do not balance all three (Gunawan and Sen, 2008; Singh and Hirsh, 1992) and only very few give any indication of how the balancing was done. Caclin *et al.* (2005) provide equations to show how loudness and duration of stimuli were balanced, but do not indicate this for pitch balancing. Balancing pitch by equating fundamental frequencies may be sufficient in many instances, but given that a dimension has been found that was correlated with F_0 when changing the spectral centroid (Marozeau and De Cheveigné, 2007), it may be important to balance stimuli for pitch as well.

In summary, several timbre properties together control the timbre of an instrument sound. In order to perform psychoacoustic tests, a synthesis method is required in which a chosen timbre parameter can be varied independently from other timbre parameters, while possible confounding cues should be eliminated by balancing the perceptual attributes loudness, pitch, and perceived duration. This study investigated (1) the use of additive synthesis to achieve control over timbre parameters, (2) to which extent loudness, pitch, and perceived duration of instrument sounds are influenced by variations in timbre parameters and which of these should be balanced for, and (3) whether pre-experiment balancing is viable. For example, some previous studies applied balancing equations derived from one set of listeners to experiments with other listeners. The question considered here is, if details of the instrument sounds to be presented (and specifically the timbre property values) were known before commencement of experiments in a timbre perception study, may loudness balancing (or balancing of the other perceptual attributes) of sounds be carried out during the stimulus preparation phase as an alternative to tedious individual balancing for all participants?

II. STIMULUS SYNTHESIS

A. Selection of recordings and timbre properties

Thirteen quasi-harmonic instrument recordings (clarinet, oboe, flute, saxophone, trumpet, tuba, French horn, trombone, bowed and plucked violin, bowed and plucked cello, and piano) of C4 (262 Hz) obtained from The University of Iowa Electronic Music Studios (Fritts, 1997) were selected to represent families of instruments (McAdams *et al.*, 1995). Instruments were characterized as being either "sustain instruments" or "decay instruments." Sustain instruments are those in which the production method is continuous; for example, the continuous bowing of a violin or cello or the application of continuous breath pressure to a trumpet, clarinet, or oboe. A decay instrument is characterized by abrupt application and removal of the production method, such as plucking the string of a violin or cello, or the hammer action against a piano string.

A set of the most salient timbre properties had to be selected to serve as the basis for instrument sound synthesis. Most multidimensional timbre studies consistently find dimensions correlating to brightness (T_b) (Grey, 1977; Marozeau *et al.*, 2003; McAdams *et al.*, 1995; Krimphoff *et al.*, 1994) and often spectral flux is also found as a dimension (Grey, 1977; McAdams *et al.*, 1995). In McAdams *et al.* (1995) and Krimphoff *et al.* (1994), one dimension also corresponds to the attack (characterized by the LRT). The influence of local spectrum variation or the shape of the spectral envelope has also been recognized as a salient attribute (Gabrielsson and Sjögren, 1971; Grey and Gordon, 1978; Gunawan and Sen, 2008; Horner *et al.*, 2004, Marozeau *et al.*, 2003).

Although spectral flux [also known as amplitude envelope coherence (McAdams *et al.*, 1999), amplitude synchronicity (Grey, 1977), and spectral evolution (Chowning, 1973; Cusack and Roberts, 2004)] have been included in timbre perception studies and instrument sound synthesis techniques, there is no clear agreement on the salience of each. Multidimensional scaling performed on similarity ratings of 16 instruments revealed one dimension related to spectral fluctuations throughout the duration of instrument tones as well as a closely related dimension of higher harmonics synchronicity (Grey, 1977). One of the timbre dimensions in the study of McAdams *et al.* (1995) corresponded to spectral flux. Caclin *et al.* (2005) found that varying spectral flux was a salient feature in dissimilarity ratings, but observed that its contribution to dissimilarity ratings decreased when co-varied with the spectral centroid (T_b) and attack time. They concluded that spectral flux was a less salient timbre feature compared to the spectral centroid and attack time.

Four timbre properties regarded as the most salient were selected to serve as the basis for the present instrument sound synthesis model. These were two spectral properties (spectral centroid, T_b, and spectral irregularity, IRR) and two temporal properties (LRT, and a property describing the remainder of the temporal envelope, called sustain/decay or SD). The selection of the first three properties was based on literature, while the temporal envelope property was selected after initial synthesis experiments. Specifically, analysis of the timbre properties T_b, IRR, LRT of the selected instrument recordings (Krimphoff *et al.*, 1994), and initial synthesis experiments to recreate the instrument sounds from the set of timbre properties showed that a temporal property describing the progression of the note (i.e., whether the note decays, for example, in the case of a piano or plucked violin string, or whether the note is sustained, for example, in the case of a flute, trumpet, or bowed violin) was required. This is the sustain/decay property SD, expanded on below. In support of the selection of SD as one of the timbre parameters in the present study, Marozeau *et al.* (2003) found a dimension correlating to the descriptor "impulsiveness," which is a dimension that very effectively separates sustained and decaying instruments into two separate perceptual groups. Other studies on musical instrument recognition recognize decay vs sustain as one of the most basic levels in the hierarchy of the categorization of musical instruments (e.g., Martin, 1999).

B. Estimation of timbre parameters from instrument recordings

Each of the four timbre parameters was determined for each of the instrument recordings. First, the harmonics of each instrument are determined by performing a fast Fourier transform on each recording and then finding the maximum value within the band $[(k-0.5)\,f_0;\,(k-0.5)\,f_0]$ where f_0 is the fundamental frequency of the recording and k is the harmonic number.

Having determined the harmonics, the spectral centroid of each instrument was evaluated as

$$T_b = \frac{\sum_{k=1}^{N} k a_k}{\sum_{k=1}^{N} a_k}, \tag{1}$$

where a_k is the amplitude of the kth harmonic and N is the total number of harmonics (Krimphoff *et al.*, 1994). T_b is obtained in harmonic rank units and $T_b \geq 1$.

More than one definition of spectral irregularity is found in literature. Although irregularity as defined by Krimphoff *et al.* (1994) provides valuable intuitive information on the deviations of harmonic amplitudes from the local spectral envelope, it is mathematically irreversible during synthesis and dependent on the number of harmonics used for the calculation. The definition by Jensen (1999b) was preferred [Eq. (2)]. The IRR of each instrument was evaluated as

$$\mathrm{IRR} = \frac{\sum_{k=1}^{N} (a_k - a_{k+1})^2}{\sum_{k=1}^{N} a_k^2}, \tag{2}$$

where a_k is the amplitude of the kth harmonic and N is the total number of harmonics. The $(N+1)$th partial is assumed to be zero (Jensen, 1999a). Irregularity is dimensionless and ranges from $0 \leq \mathrm{IRR} \leq 2$.

Using the method of Jensen (1999a), LRT is determined from an estimate of the attack time of the instrument recording. This method defines the attack as the point in time where the highest positive slope of a temporal envelope in which slight amplitude variations have been smoothed is observed. The start-of-attack (soa) and end-of-attack (eoa) is found in milliseconds (Jensen, 1999a, pp. 56–58), which can then be used to determine the LRT of the instrument recording as

$$\mathrm{LRT} = \log(\mathrm{eoa} - \mathrm{soa}). \tag{3}$$

Finally, the second part of the temporal envelope (after the initial attack segment) is defined here as the SD segment of the note. As LRT provides no information about the remainder of the tone after the initial attack (Krimphoff *et al.*, 1994), Jensen's temporal envelope model (Jensen, 1999a,b) that also describes the SD and release segments in addition to the attack was adapted. The remainder of the temporal envelope after the eoa was modeled by a one-parameter curve fit that combined Jensen's SD and release segments. Using the Levenberg-Marquardt curve-fitting algorithm, the equation

$$\hat{y}(\tau) = y_0 + (y_1 - y_0)\frac{e^{\mathrm{SD}\,\tau} - 1}{e^{\tau} - 1}, \tag{4}$$

was fitted through the SD segment to obtain an estimate value of SD. $y(\tau)$ is the amplitude of the temporal envelope segment as a function of normalized time ($0 <= \tau <= 1$), while y_0 and y_1, respectively, indicate the amplitudes at the start and end of the temporal envelope segment. For the present study, $y_0 = 1$ and $y_1 = 0$. When $\mathrm{SD} > 0$, the instrument sound is sustained and for $\mathrm{SD} < 0$, the instrument sound is decaying faster than a linear decay. As $\mathrm{SD} \to 0$, the curve approaches a linear decay.

The four parameters (T_b, IRR, LRT, and SD) for each instrument determined from the instrument recordings are summarized in Table I.

C. Synthesis of instrument sounds from timbre parameters

The spectrum of each instrument sound was recreated using additive synthesis [Eq. (5)]. Additive synthesis simply adds sinusoids of amplitudes a_k at harmonic frequencies $f_k = kf_0$ where f_0 is the fundamental frequency and k is the kth harmonic

$$s(t) = \sum_{k=1}^{N} a_k \sin(2\pi k f_0 t). \tag{5}$$

To recreate a sound with a target spectral centroid, the harmonic amplitudes a_k need to be known. Following Jensen (1999b), an intermediate variable B is defined such that

$$a_k = B^{-k}, \tag{6}$$

TABLE I. Timbre parameter values for 13 instruments. T_b is the spectral centroid and is measured in harmonic rank units, IRR is the spectral irregularity (dimensionless), LRT is the LRT measured in log(s), and SD is the dimensionless parameter of the SD temporal envelope segment.

Instrument name	T_b	IRR	LRT	SD
Clarinet	3.00	1.16	2.07	17.0
French horn	2.41	0.192	1.56	18.2
Oboe	5.23	0.603	1.66	12.3
Violin (bowed)	5.25	0.568	2.50	6.52
Violin (plucked)	2.58	0.562	1.22	−19.6
Flute	3.70	0.133	2.09	8.04
Piano	2.42	0.297	1.27	−7.49
Trumpet	5.41	0.185	1.53	14.8
Tuba	2.57	0.329	1.55	4.28
Cello (bowed)	6.65	0.991	2.74	2.16
Cello (plucked)	1.88	0.920	1.18	−12.9
Saxophone	3.46	0.330	2.16	2.20
Trombone	4.18	0.227	2.00	8.69

where a_k is the amplitude of the kth harmonic. With B defined as above, it is related to the spectral centroid T_b in the following way [from Eq. (1)]:

$$B = \frac{T_b}{T_b - 1}, \tag{7}$$

so that, for a given target centroid, the amplitudes of the synthesized harmonics may be determined. The spectrum is then synthesized as described by Eq. (5). The number of harmonics was limited to $N = 30$ in the present implementation.

To synthesize a sound with a given spectral irregularity, odd harmonics need to be smaller or larger than even harmonics by a multiplication factor, x, in Eq. (8) below

$$\begin{aligned} a_k &= xB^{-k} \quad \text{for} \quad k = 1, 3, 5 \ldots \\ a_k &= B^{-k} \quad \text{for} \quad k = 2, 4, 6 \ldots . \end{aligned} \tag{8}$$

As the spectral centroid and irregularity are interdependent, solutions for B and x must be obtained by substituting the desired centroid and irregularity values of the sound to be synthesized and solving the system of Eqs. (9) and (10). These equations follow when Eq. (8) is substituted in Eqs. (1) and (2) with $N \to \infty$

$$T_b = \frac{xB^3 + 2B^2 + xB}{xB^3 + B^2 - Bx - 1}, \tag{9}$$

$$\mathrm{IRR} = \frac{2B^2 - 2xB^3 + x^2B^4 + x^2 - 2xB}{x^2B^4 + B^2}. \tag{10}$$

Valid solutions for x and B are then used in Eq. (8) to obtain the desired harmonic amplitudes, and spectral synthesis is achieved through Eq. (5). Different values of T_b and IRR yield different root-mean-square (rms) values for each synthesized sound. Therefore, the intensities of these were equalized for intensity by scaling the signals to have equal rms values.

The temporal envelope of each instrument sound was recreated using a linear segment rising from zero to maximum amplitude for the duration $\mathrm{RT} = 10^{\mathrm{LRT}}$ in milliseconds (ms). The SD segment was recreated using Eq. (4). The two segments (attack and SD) were concatenated to realize the complete temporal envelope. The temporal envelope was recreated be of 2 s duration.

Finally, the temporal envelope is multiplied by the periodic signal, $s(t)$ resulting from the additive synthesis. Each instrument sound was synthesized in this fashion, giving complete control during synthesis over each timbre parameter [for spectral parameters, within the bounds set by Eqs. (9) and (10)]. When these synthesized sounds are used in timbre perception studies, they should be balanced to be of equal loudness, pitch, and perceived duration. This is considered in Sec. III.

III. LOUDNESS, PITCH, AND DURATION BALANCING

Generally, balancing sounds for equal loudness, pitch, and perceived duration is a necessary step for discrimination or dissimilarity tasks (Grey, 1975). Perceptual data was obtained to determine when balancing was necessary and to determine whether data would allow trends (e.g., loudness changes when the spectral centroid is varied) to be quantified so that balancing need not be done for every listener individually (i.e., pre-experiment balancing). If this were possible, it would eliminate tedious balancing tasks in experiments. Data were measured by means of balancing tasks for the following conditions: Loudness changes with variation in spectral centroid, irregularity, rise-time, and SD time; perceived duration changes as rise-time and SD times varied; pitch changes as spectral centroid and irregularity varied. In each condition, the specific timbre parameter tested was varied across a range of values, expanded on below.

A. Listeners

Six listeners (4 females and 2 males aged between 21 and 28 yrs with an average age of 25) with normal hearing [pure tone thresholds $\leq$20 dB hearing level (HL) for 250, 500, 1000, 2000, 4000, and 8000 Hz] were used to obtain perceptual data. The total experiment duration was between 5 and 7 h for each listener. Listeners were compensated for their time.

B. Characterization of the SD parameter

1. Procedure and stimuli

When varying the timbre parameters across a range of values it was necessary to efficiently distribute levels at which the particular parameter was presented. The relationship between SD and loudness was hypothesized to be of logarithmic nature; this was tested in a first experiment.

Pure tones of 262 Hz with a total duration of 2 s, a sampling rate of $f_s = 44\,100$ Hz, and an LRT of 1 were presented. Tone decay was varied with logarithmically spaced SD values with $\mathrm{SD} = [-10^x, 0, \text{and } 10^x]$ and $x = [-1.8, -1.6, -1.4, \ldots, 1.4, 1.6, 1.8]$. All tones were presented at the 75% loudness level of the subjects' individual loudness growth curves, presentation levels varying between 68.2 dB sound pressure level (SPL) and 76 dB SPL.

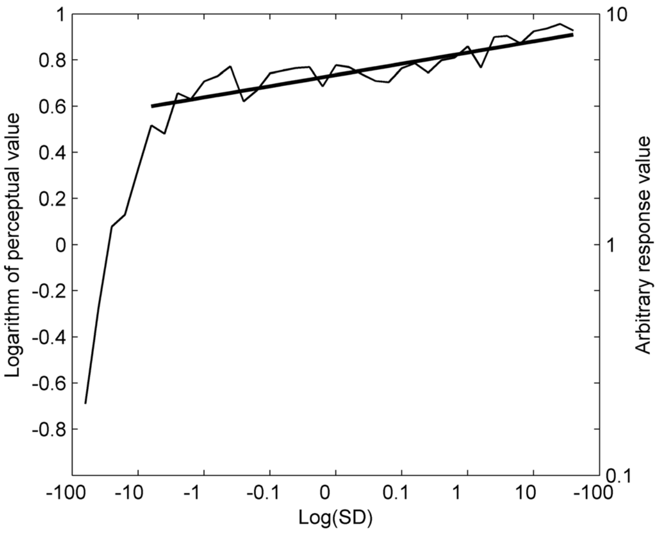

FIG. 1. The logarithm of the average of the perceptual responses (similarity to the extreme values of SD) is shown to be linearly related to logarithmically spaced SD-values across most of the range of SD. Note that SD is a dimensionless parameter.

Experiments were conducted in a double-walled sound booth complying with ANSI S3.1–1999. Tone presentation was controlled by software on a personal computer. Tones were presented through a KEF Q30 (GP Acoustics (UK) Ltd, Kent, UK) loudspeaker via an M-Audio Fast-track Pro (inMusic Brands, Inc, Cumberland, RI) external soundcard. Listener responses were recorded automatically by the software.

In each repetition, three tones were presented with the first and last tones representing the extremes of the range ($SD = -10^{1.8}$ and $SD = 10^{1.8}$). The middle tone corresponded to a tone within the range and listeners had to decide to which extent this tone was perceived to be similar to either of the extreme tones by indicating this on an analog scale on the computer interface. Each of the 38 SD values in the range was presented once.

2. Results

Figure 1 plots the average values of six listeners on logarithmic axes. Figure 1 indicates that the logarithm of the perception of SD was linearly related to logarithmically spaced values of SD across most of the range over which SD was varied, confirming the initial suspicion. The linear regression coefficient for the segment of the curve above $SD = -6$ was 0.048. Based on this outcome, logarithmically spaced values of SD were used when varying this timbre parameter in further experiments.

C. Balancing of loudness for variations in timbre parameters

1. Procedure and stimuli

The values for the four timbre parameters that were determined for each instrument were used to synthesize tones with the additive synthesis model. Stimuli in the loudness balancing experiment were synthesized by varying the value of one timbre parameter [T_b, IRR, LRT, or log(SD)] in eight increments, while the other three parameters were kept constant. Listeners were presented a reference synthesized instrument tone (using the parameters in Table I) and had to match the loudness of a test tone in which the timbre parameter under consideration was varied, to the reference using a slider bar on a computer interface. The slider bar controlled the test tone's intensity and could be adjusted 6 dB SPL in both directions from the intensity of the reference tone. Subjects were allowed to listen to the tone pair as many times as desired and could save their response when satisfied that the two tones were equal in loudness. Stimuli were presented in random order. This was repeated for all 13 instruments and for each of the 4 timbre parameters, amounting to a total of 416 loudness balancing tasks for each listener. Software controlled the progress of the experiment and saved listener responses.

The timbre parameter ranges used for these experiments are summarized in Table II. The irregularity column and lower spectral centroid value column contain the extreme values that allow simultaneous solution of Eqs. (9) and (10). The higher spectral centroid limit was set to twice the reference tone spectral centroid value. Extreme LRT values used during balancing of perceived duration were respectively half and twice that of the original LRT of the instrument, except when the higher extreme exceeded 3.1, in which case the upper extreme was set to 3.1.

TABLE II. Boundaries of the values used for the balancing tasks.

	T_b		IRR		LRT (except duration)		LRT (duration)		SD	
Clarinet	2.35	6.00	0.11	1.25	1	3	1.03	3.1	4.25	68.00
Horn	2.11	4.82	0.16	1.17	1	3	0.78	3.1	4.55	72.84
Oboe	1.55	10.45	0.04	1.46	1	3	0.83	3.1	3.08	49.24
Bowed violin	1.54	10.50	0.04	1.46	1	3	1.25	3.1	1.63	26.06
Plucked violin	1.53	5.15	0.14	1.19	1	3	0.61	2.44	−4.90	−78.34
Flute	2.59	7.40	0.07	1.17	1	3	1.05	3.1	2.01	32.17
Piano	1.62	4.83	0.16	1.17	1	3	0.63	2.54	1.87	29.98
Trumpet	2.15	10.81	0.04	1.47	1	3	0.77	3.07	3.70	59.28
Tuba	1.51	5.14	0.14	1.19	1	3	0.78	3.1	1.07	17.12
Bowed cello	1.67	13.30	0.03	1.54	1	3	1.37	3.1	0.54	8.63
Plucked cello	1.65	3.76	0.24	1.09	1	3	0.59	2.35	−3.22	−51.46
Saxophone	1.51	6.91	0.08	1.30	1	3	1.08	3.1	0.55	8.80
Trombone	1.90	8.36	0.06	1.37	1	3	1.00	3.1	2.17	34.78

2. Results

Figures 2–5 document the intensity changes required for equal loudness of the test and reference tones for spectral centroid, irregularity, LRT, and SD, respectively. Each data point is the average across six listeners, and the error bars show the standard deviation across listeners.

Considering T_b (Fig. 2), it is clear that changes in spectral centroid result in loudness changes. Intensity had to be adjusted downwards for all instruments as the spectral centroid shifted toward higher frequencies, indicating that loudness of instrument sounds increased with increased spectral locus frequency. This was confirmed by an Ancova (analysis of covariance) in which intensity adjustments were compared across the 13 instruments and across all listeners. This indicated a main effect of spectral centroid, $p < 0.05$. The Ancova analysis was also used to test whether the intensity adjustment slopes were equal across instruments. If true

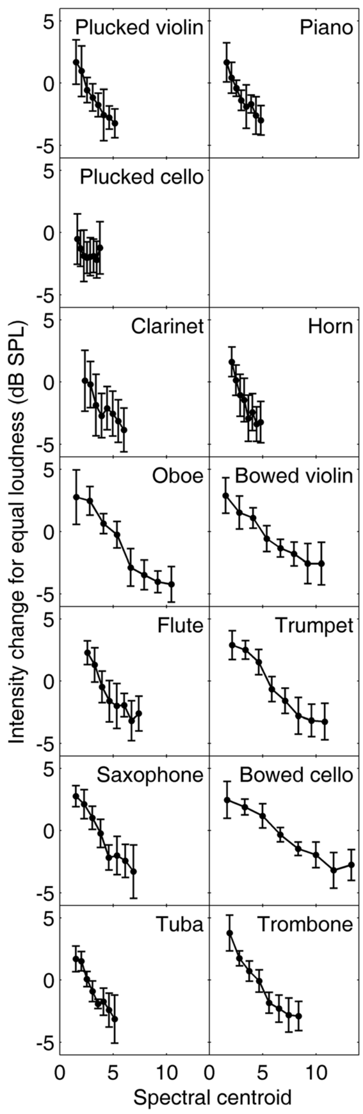

FIG. 2. Loudness balancing data for variation in spectral centroid. The average response and standard deviations across listeners are shown, with error bars indicating one standard deviation. The first three panels show data for the decay instruments, while the remaining panels show the sustain instrument data. Note that the spectral centroid parameter is dimensionless.

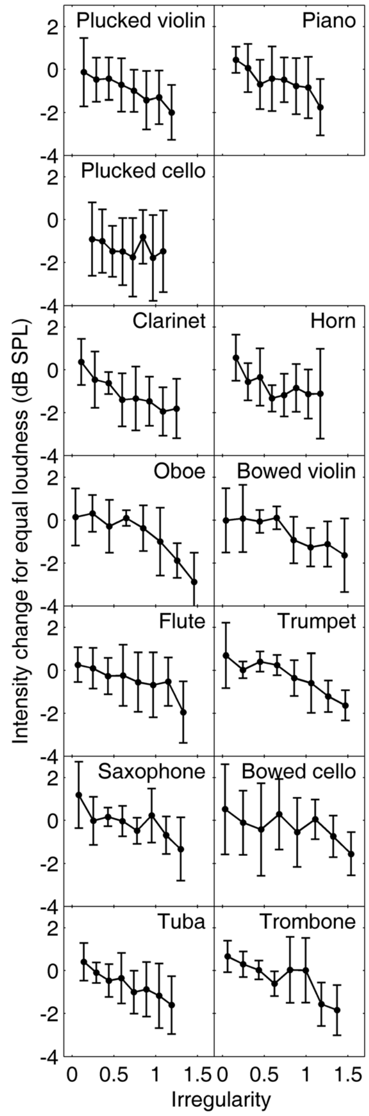

FIG. 3. Loudness balancing data for variation in irregularity. The average response and standard deviations across listeners are shown. Error bars indicate one standard deviation. As in Fig. 2, the decay instrument data are shown in the first three panels. Irregularity is dimensionless.

and if the spectral centroid of a particular instrument sound were known, this would allow pre-experiment balancing. Significant interaction [$F(12, 598) = 16.12$, $p < 0.05$] between T_b and instrument, however, indicated that intensity adjustment slopes differed for different instruments. This means that it would be necessary to perform loudness balancing for each instrument individually when varying the spectral centroid to elicit brightness changes. Also, the range of intensity changes required for equal loudness may differ by 5 dB across listeners, which is larger than typical intensity discrimination thresholds, so that balancing equations derived from one set of listeners may not be applicable to another.

Loudness balancing data for IRR (Fig. 3) shows that irregularity changes result in loudness changes. As for the spectral centroid, intensity had to be adjusted downwards for all instruments as irregularity increased, indicating that loudness increased with increasing irregularity. Changes in loudness resulting from varying irregularity are generally smaller than when the spectral centroid is varied. Interaction

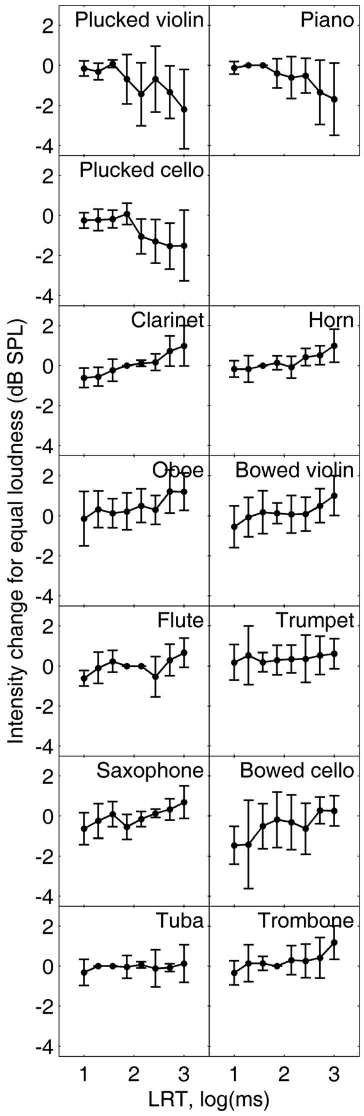

FIG. 4. Loudness balancing data for variation in LRT. The average responses and standard deviations across listeners are shown for each instrument with the first three panels corresponding to the decay instruments and the subsequent panels corresponding to sustain instruments.

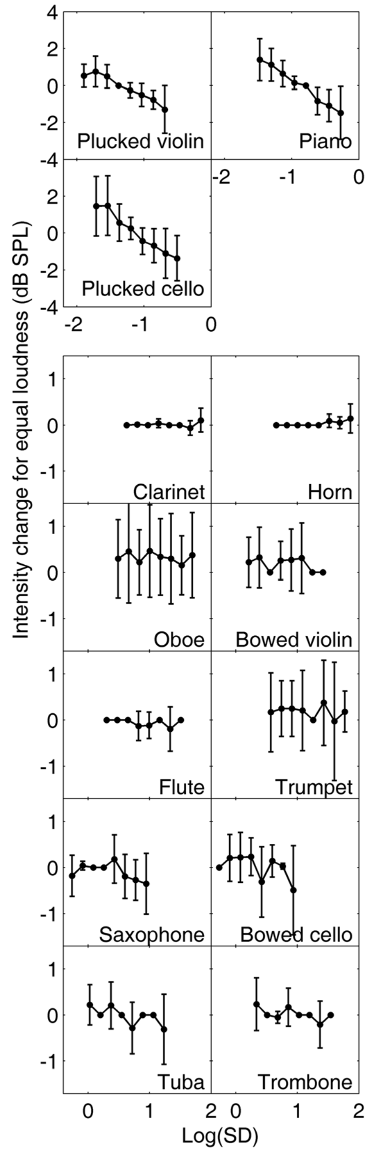

FIG. 5. Loudness balancing data for variation in SD (a dimensionless parameter). Average responses and standard deviations across listeners are shown for each instrument as values of SD were varied. The first three panels correspond to the decay instruments, while the subsequent panels correspond to sustain instruments.

between IRR and instrument was not significant [$F(12, 598) = 1.12$, $p > 0.05$], indicating that slopes of the psychometric function did not differ significantly across instruments. This means that it may be viable to use the same function for all 13 instruments to compensate for loudness differences when IRR is varied. Required adjustments to compensate for loudness differences resulting from variation in IRR are generally smaller than 3 dB SPL. The variance across listeners is small for some instruments and up to around 4 dB for others, which exceeds typical intensity discrimination thresholds, so that balancing equations derived from one set of listeners may not be generally applicable.

For decay instruments (the first three panels in Fig. 4, i.e., plucked violin, piano, and plucked cello), slopes of the psychometric function relating LRT to loudness do not differ significantly [$F(2, 138) = 0.36$, $p > 0.05$]. However, the slopes for the sustain instruments (all the other panels in Fig. 4) do differ significantly [$F(9, 460) = 3.79$, $p < 0.05$]. This suggests that it may be possible to compensate for the changes in loudness resulting from LRT of the three decay instruments using the same adjustments for all three instruments, while the same adjustment equation cannot be used for all ten sustain instruments. Required adjustments to compensate for loudness differences when LRT is varied are generally smaller than 2 dB SPL.

Figure 5 shows the intensity adjustments required for decay (first three panels) and sustain (remainder of the panels) instruments to achieve equal loudness when SD is varied. Although psychometric function slopes differ significantly for sustain instruments [$F(9, 460) = 3.71$, $p < 0.05$], adjustments were slight (generally below 1 dB SPL, with a maximum adjustment of 1.9 dB SPL in two instances out of a total of 480 adjustments made) and the adjustment slope is close to 0. The psychometric function slopes of the three decay instruments also differed significantly [$F(2, 138) = 18.6$, $p < 0.05$].

The results of the sustain instruments perhaps need further clarification. The instances where the standard deviations are zero indicate that no changes to loudness levels were made by any listener and that they decided that the two originally presented tones were equal in loudness. Only 73 sound level adjustments were made for a total of 480 pairs of stimuli presented, all by listeners 1 and 2.

Figure 6 gives a general view of these results. Although psychometric function slopes of different instruments differ in some conditions, intensity adjustments required for equal loudness for all instruments and all listeners were pooled in this figure. Linear regression lines are shown in each of the panels of Fig. 6. To obtain these, the data set of each instrument was displaced up or down along the intensity axis. Where the spectral centroid, irregularity, and SD-values were not equal across instruments (Table II), instruments were grouped together in intervals. For example in Fig. 6(a), the data point at a centroid value of $T_b = 2$ is the average intensity adjustment for all participants' responses across all instruments for the interval $T_b = [1.5, 2.5]$. The error bars indicate the standard deviations of these responses. The range of LRT values used for loudness balancing was identical across instruments.

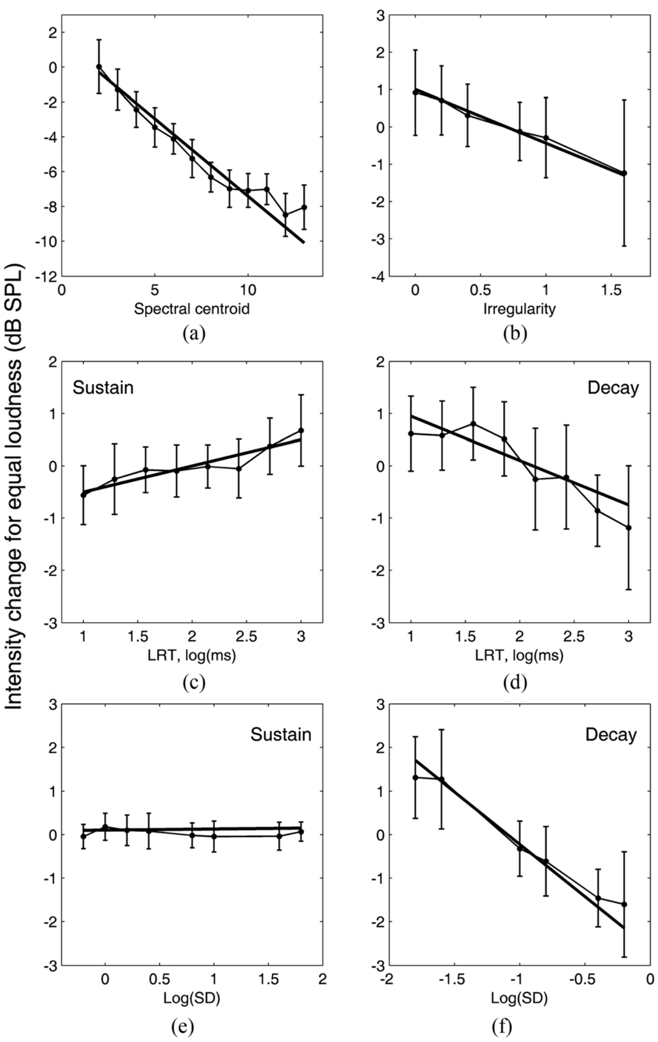

FIG. 6. Loudness balancing data for the four timbre parameters averaged across instruments and listeners, with error bars indicating one standard deviation. For LRT and SD, data for sustain and decay instruments are shown on separate panels. Linear regression lines are shown in all panels.

Table III contains the gradient values from the linear regression fits to the data. For all timbre parameters, except SD for the sustain instruments ($SD_{sustain}$), regression lines fit the data reasonably well. With the exception of $SD_{sustain}$, standard deviations calculated from the error variances (Table III) are small compared to the differences in gradients.

Figure 6 and Table III suggest that it would be reasonable to balance loudness of stimuli when one of the four timbre parameters are varied by adjusting the intensity of the stimulus according to the linear regression slopes of Table III. Based on Caclin *et al.* (2005), Eq. (11) may be used to perform balancing tasks

$$A(T_a) = A(T_r) \times 10^{G\times(\mathrm{TP}(T_r)-\mathrm{TP}(T_a))/20}, \tag{11}$$

with T_r and T_a the reference and intensity-adjusted tones, respectively, TP the timbre parameter being varied, and G the gradient indicated in Table III.

D. Balancing of perceived duration for variations in timbre parameters

1. Procedure and stimuli

Stimuli were varied similarly to the loudness balancing procedure. Assuming that the two spectral timbre parameters (T_b and IRR) do not influence perceived duration, eight equally spaced values for LRT and log(SD) for each instrument were presented using the limits indicated in Table II. These two timbre parameters were varied in separate experiments. An adjustment of the duration of decay type instrument sounds was not possible due to their nature. While a performer can produce short or long notes for sustained instruments (e.g., violin or clarinet), a struck or plucked string (e.g., piano or plucked violin) with a particular SD always has similar duration. An increase in the duration of the note (e.g., by depressing the pedal that lifts the piano mutes from the strings) yields a tone of longer duration, but also changes the decay rate of the note, which changes SD. Therefore, the ten sustain instruments were used for the perceived duration balancing task, but decay instruments were not included in this task.

Listeners had to match the duration of a test tone of which one temporal parameter (SD or LRT) was varied to a reference tone by using a slider bar on a computer interface. The duration of the reference tone was always 2 s and the test tone duration could be varied from 1.5 to 2.5 s by

TABLE III. Gradients, R^2 values, and variances found by linear regression fits to the loudness response data.

Parameter	Gradient [95% confidence intervals]	R^2 value	Error variance
T_b	0.873 [0.806, 0.940]	0.868	0.7544
IRR	1.46 [1.29, 1.64]	0.727	0.132
LRT_{decay}	−0.498 [−0.588, −0.409]	0.613	0.0688
$LRT_{sustain}$	0.984 [0.743, 1.23]	0.764	0.139
SD_{decay}	2.34 [1.83, 2.86]	0.804	0.269
$SD_{sustain}$	0.0876 [0.0248, 0.150]	0.090	0.0183

inserting a 0.5 s segment into, or deleting a 0.5 s segment from, the sustained part of the tone. Listeners were allowed to listen to the tone pair as many times as desired and could save a response when satisfied that the two tones were of equal perceived duration. The task was repeated for each of the ten instruments tested and stimuli were presented in random order for each of the two timbre parameters.

2. Results

Figure 7 shows the respective duration adjustments as the LRT and SD values changed. Adjustments were generally small. One-way repeated measures analysis of variance indicated no significant differences in responses for parameter changes. Linear regression analyses indicated large standard deviations in adjustments for both LRT and SD. These data suggest that balancing of perceived duration is not required for variations in SD and LRT, at least not for stimuli of the duration used (2 s).

E. Balancing of pitch for variations in timbre parameters

1. Procedure and stimuli

As it was assumed that the temporal timbre parameters (SD and LRT) do not affect pitch of the stimuli, only the effects on pitch of variations in T_b and IRR were tested. Stimuli were varied in eight equally spaced increments of T_b and IRR. The same limits used for the loudness balancing task (Table II) were used for the pitch balancing task. Subjects were presented with a reference tone and had to match the pitch of a test tone in which one of the two spectral timbre parameters was varied to the pitch of the reference tone using a slider bar on the computer interface. The pitch balancing task differed somewhat from the loudness and duration balancing tasks as the starting frequency of the test tone was not equal to that of the reference tone. The reference tone was middle C ($\approx$262 Hz) and the test tone's frequency was selected randomly to be within a 262 Hz $\pm$ 7.5 Hz band. The slider bar could be used to vary the test tone frequency within $\pm$10 Hz of the initial test tone frequency. Listeners could repeat the test and reference tones as many times as desired and could save a response when satisfied that the two tones were equal in pitch. The experiment was repeated for each of the 13 instruments and stimuli were presented in random order for each of the 2 timbre parameters tested.

2. Results

Two subjects performed the task with ease, while the other four subjects anecdotally stated that the task was difficult and that they were unsure of their responses and their ability to carry out the instructions. The data supported their verbal reports. The data from the subjects who were able to perform the task had low standard deviations (0.55 and 0.80 for spectral centroid and 0.49 and 0.72 for irregularity) compared to the other subjects (standard deviations larger than 4.5 for spectral centroid and 3.8 for irregularity).

Figure 8 shows the average responses of the two listeners who were able to perform the task. Ninety-five percent of their responses fell respectively within a $\pm$0.8 Hz and a $\pm$1.5 Hz band for the spectral centroid and within a $\pm$1 Hz and a $\pm$1.5 Hz band for irregularity. These values are comparable to perceptual discrimination in pure tone frequencies (between approximately 0.5% and 1% for sound levels used in this experiment) (Shower and Biddulph, 1931; Wier *et al.*, 1977). This observation suggests that, for the parameter ranges tested here, pitch balancing may not be necessary.

IV. DISCUSSION

A. Stimulus synthesis

The present study proposes that an additive synthesis model may be of value in timbre perception studies. Of particular benefit is the separation of spectral and temporal properties and the ability to vary timbre parameters in a controlled manner.

While synthesis models generally include attack time as the principal temporal timbre property (Krimphoff *et al.*, 1994), the present work included two temporal properties that characterize the entirety of the temporal envelope (Jensen, 1999a), rather than only the attack.

The analysis and re-synthesis of tones with a specific brightness is established in literature (Caclin *et al.*, 2005; Jensen, 1999b, 2001; Krimphoff *et al.*, 1994). However, methods for synthesis of tones with a particular irregularity are not consistent across studies. Caclin *et al.* (2005), Jensen

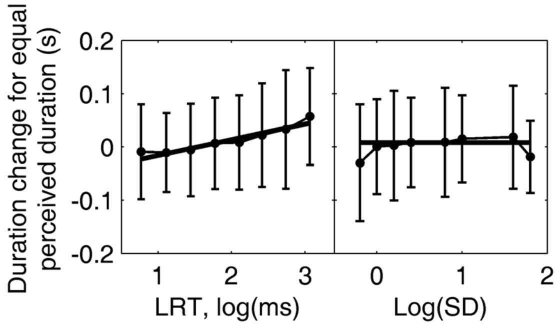

FIG. 7. Required adjustment in duration for equal perceived duration when LRT (left panel) or SD (right panel) is varied for sustain instruments. Data are averaged across listeners and instruments. Linear regression lines are shown in both panels.

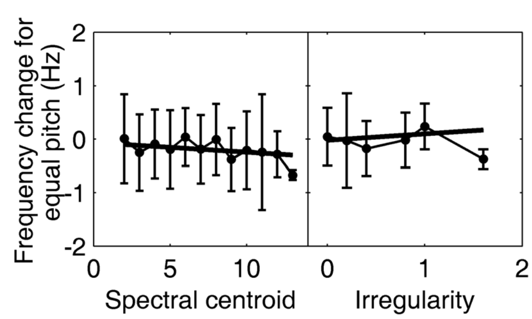

FIG. 8. Required adjustment in frequency to obtain equal pitch when the spectral centroid (left panel) and irregularity (right panel) are varied. Only two listeners were able to perform the task and the data from these two listeners are averaged across listeners and instruments. Linear regression lines are shown in both panels.

(1999b), Krimphoff *et al.* (1994), McAdams *et al.* (1999), and McAdams *et al.* (1995) all define or apply irregularity differently. In addition, it should be noted that changes in irregularity and spectral centroid are not independent. Consequently, Eqs. (9) and (10) are of value to simultaneously achieve the desired spectral centroid and irregularity as defined by Eqs. (1) and (2). It is important to note that although spectral centroid and irregularity can both be specified, and can be varied independently through Eqs. (9) and (10), it should not be assumed that their perceptual correlates (brightness and hollowness) are perceptually orthogonal. Perceptually, one may still influence the other.

B. The influence of changes in timbre parameters on loudness, pitch, and perceived duration

1. Loudness

One of the objectives of the present study was to consider which timbre parameters influence loudness and would therefore require loudness balancing in studies of timbre perception. The data presented show that changes in all four timbre parameters investigated (spectral centroid, irregularity, attack, and decay rate) affect the loudness of an instrument tone presented at a given intensity, although not all to the same extent. The data suggest that balancing is probably not required for all four timbre parameters, at least not for the ranges of these parameters tested. Average intensity changes to achieve equal loudness are slight for some timbre parameters, so that balancing will not be necessary, while other timbre parameters require balancing. Where intensity changes required for equal loudness are within intensity discrimination thresholds for NH listeners (around 1 dB), balancing is probably unnecessary.

Spectral centroid variation (and therefore brightness) strongly influenced loudness, as may be expected from literature. Loudness is known to be a function of the spectral composition of a sound. For example, existing models predict loudness to be a function of the bandwidth of a sound. Spectral loudness summation is a well-known characteristic of the auditory system (e.g., Zwicker and Fastl, 1999). Scharf (1962) showed that the loudness of a complex sound is related to the summation of the loudnesses of the different critical bands within which the spectral components of this sound falls. While bandwidth of a sound does not affect loudness if all the spectral components fall within one critical band, loudness increases significantly when the bandwidth is wider than a critical band, as seen, for example, in the data of Verhey and Kollmeier (2002).

Bulen (1995) documented data that relate loudness of an instrument to brightness; data were shown for trombone, saxophone, and French horn. Bulen found that log brightness varies approximately quadratically with log loudness. While Bulen measured brightness as a function of loudness, the measurements in the present data measured loudness as a function of brightness at a fixed sound intensity. As may be expected though, the relationship between brightness and loudness corresponds to that found by Bulen, showing that loudness increases as brightness increases (i.e., as spectral centroid shifts to higher frequencies). It may also be noted that (Fig. 2) the slope of the loudness increase, decreases for spectral centroids at higher frequencies (higher brightness values). This may be explained by loudness models (e.g., Zwicker and Fastl, 1999). These calculate the contribution of individual harmonics using the critical band rate scale, so that lower harmonics fall in different critical bands, but higher harmonics may fall into the same critical band. This means that higher frequency components in the spectrum contribute less to loudness, as is also seen in the data of Oberfeld *et al.* (2012). So, as the spectral centroid shifts toward higher frequencies, there will be a point at which the slope of loudness increase will flatten, as observed in the present data.

As spectral centroid variation strongly influenced loudness, balancing of loudness of stimuli is required when the spectral centroid is manipulated or instruments with different spectral centroid values are compared. Similarly, balancing of loudness is required when tones have different irregularity values, different rise-times, and different decay times. The latter is true for decay instruments, but the minor influence on loudness when varying the SD parameter for sustain instruments obviates the need for loudness balancing in this case.

Balancing has not always been done (or reported) in published timbre perception studies. Emiroglu and Kollmeier (2008) mention that differences in duration were eliminated by removing the attack segment of recordings, but provide no explanation of participants or methods used to balance loudness and pitch. Similarly, Gabrielsson and Sjögren (1971) do not indicate whether tones were balanced in loudness, pitch, or duration. Some studies indicate specific details of loudness and/or duration adjustments, but do not mention specific adjustments for pitch (McAdams *et al.*, 1995; Caclin *et al.*, 2005).

A second objective was to consider if pre-experiment balancing was feasible. That is, may loudness, pitch, or duration balancing of sounds be carried out during the stimulus preparation phase, and may balancing equations obtained for one set of listeners be used for another? The present data suggest that this should not be standard practice, although it may still be feasible under certain conditions. Considering Table III, the goodness of fit of the regressions appears to support the use of Eq. (11) for intensity adjustments to obtain equal loudness. However, as shown, not all instruments have the same psychometric function slope, so that particular care has to be taken when deciding to perform pre-experiment balancing. If the expected changes in loudness corresponding to timbre parameter variation were well characterized and were known to be similar across listeners and the particular set of instruments, this may be possible. The present data suggest that it may be possible to use the same loudness adjustment function determined for irregularity changes across instruments, but not across listeners. The data indicate, however, that loudness adjustments to compensate for brightness changes should not be generalized either across listeners or across instruments.

However, examples of pre-experiment balancing do occur in literature. For example, Grey (1977), Gunawan and Sen (2008), Marozeau *et al.* (2003), and McAdams *et al.* (1995)

made use of pilot groups to balance loudness, pitch, and duration, while Caclin *et al.* (2005) developed equations to balance duration and loudness from the average responses of a pilot group of eight listeners.

2. Pitch

Although it is known that the pitch changes as timbre changes, timbre parameter adjustments required for pitch changes are usually large. Singh and Hirsh (1992) found that changes in spectral locus (T_b) often influence the perceived pitch. Russo and Thompson (2005) found a strong influence of large spectral centroid changes on perceived interval size. The spectral locus changes in these studies were relatively large, at least one unit of the spectral centroid for every trial. Vurma and Ross (2007) found that tuning tasks for piano, oboe, and voice yielded a range of "in-tune" ratings near the fundamental frequency. Frequency ranges where subjects (musically trained vocalists) rated two tones as being in tune 75% of the time varied between a few cents to as many as 50 cents. At 220 Hz, 50 cents relates to a 6.4 Hz band wherein tones were judged to be in tune. It appears then that larger spectral parameter adjustments than those used in the present study are required for pitch changes to be observed. This suggests that pitch balancing is not required for the timbre parameter ranges used in the present study.

3. Perceived duration

Data presented suggest that balancing of perceived duration may not be required for variations in SD and LRT within the ranges of these parameters used in the present study. Due to the relatively long duration of the tones used, changes in duration that may have been perceptible with shorter stimuli may have been masked. Abel (1972) found that the jnd in duration, ΔT, between two tones of length T and $T + \Delta T$ increased as T increased. At $T = 1$ s, the jnd was around 50 ms. To compensate for perceived changes in duration when LRT was varied, Caclin *et al.* (2005) adjusted the total stimulus duration for their stimuli that ranged between 615 and 800 ms. Results of the present balancing task suggest that duration will probably not be a confounding cue in timbre perception studies that use longer stimulus durations than these.

V. CONCLUSIONS

(1) Additive synthesis provides a method for instrument sound synthesis in which control is gained over timbre parameters. Table II documents values of four timbre parameters (spectral centroid, irregularity, rise-time, and SD time) that may be used to synthesize 13 instruments that span the timbre space defined by these parameters. Also, although the spectral centroid and irregularity co-vary, values of these parameters may be set explicitly using the equations provided. The synthesis model may serve as the basis for more complex timbre models that may include other timbre properties.

(2) Data from the present study suggest that loudness balancing is required when varying any of the four timbre parameters, with one exception. Variation in the SD parameter does not appear to have a significant influence on loudness. The data also suggest that balancing of pitch and balancing of perceived duration is not necessary for the ranges of the timbre parameters and total stimulus duration used.

ACKNOWLEDGMENTS

This work was supported by the NRF (National Research Foundation) of South Africa.

Abel, S. M. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.

ANSI (**1994**). ANSI S1.1-1994 (R 1999). *American National Standard Acoustical Terminology* (American National Standards Inst., New York).

Bulen, J. C. (**1995**). "Brightness measures of trombone timbre," Ph.D. Dissertation, University of Washington, Seattle, WA, pp. 1–192.

Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (**2005**). "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," J. Acoust. Soc. Am. **118**, 417–482.

Chowning, J. (**1973**). "The synthesis of complex audio spectra by means of frequency modulation," J. Audio. Eng. Soc. **21**, 526–534.

Cusack, R., and Roberts, B. (**2004**). "Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation," Hear. Res. **193**, 95–104.

De Poli, G. (**1983**). "A tutorial on digital sound synthesis," Comput. Music J. **7**, 8–26.

Emiroglu, S. S. (**2007**). "Timbre perception and object separation with normal and impaired hearing," Ph.D. Dissertation, Carl von Ossietzky University, Germany, pp. 1–137.

Emiroglu, S. S., and Kollmeier, B. (**2008**). "Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions," Brain Res. **1220**, 199–207.

Fletcher, N. H., and Rossing, T. D. (**1999**). *The Physics of Musical Instruments* (Springer-Verlag, New York), pp. 1–759.

Fritts, L. (**1997**). *The University of Iowa Electronic Music Studios* [Online]. Available from: http://theremin.music.uiowa.edu/index.html (Last viewed March 21, 2010).

Gabrielsson, A., and Sjögren, H. (**1971**). "Detection of amplitude distortion in flute and clarinet spectra," J. Acoust. Soc. Am. **52**, 471–483.

Galvin, J. J., Fu, Q., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**, 302–319.

Galvin, J. J., Fu, Q., and Oba, S. I. (**2009**). "Effect of competing instrument on melodic contour identification by cochlear implant users," J. Acoust. Soc. Am. **125**, EL98–EL103.

Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S., and Debus, B. (**1998**). "Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults," J. Am. Acad. Audiol **9**, 1–19.

Grey, J. M. (**1975**). "An exploration of musical timbre," Ph.D. thesis, Stanford University, California, pp. 21–23 and 41–57.

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Grey, J. M., and Gordon, J. W. (**1978**). "Perceptual effects of spectral modifications on musical timbres," J. Acoust. Soc. Am. **63**, 1493–1500.

Gunawan, D., and Sen, D. (**2008**). "Spectral envelope sensitivity of musical sounds," J. Acoust. Soc. Am. **123**, 500–506.

Horner, A., Beauchamp, J., and So, R. (**2004**). "Detection of random alterations to time-varying musical instruments," J. Acoust. Soc. Am. **116**, 1800–1810.

Jackendoff, R., and Lerdahl, F. (**2006**). "The capacity for music: What is it, and what's special about it?," Cognition **100**, 33–72.

Jensen, K. (**1999a**). "Envelope model of isolated musical sounds," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects*, Trondheim, Norway, pp. W99-91–W99-94.

Jensen, K. (**1999b**). "Timbre models of musical sounds," Ph.D. Dissertation, University of Copenhagen, Denmark, pp. 51–98.

Jensen, K. (**2001**). "The timbre model," *Workshop on Current Research Direction in Computer Music*, Barcelona, pp. 88–94 and 174–186.

Kong, Y., Mullangi, A., Marozeau, J., and Epstein, M. (**2011**). "Temporal and spectral cues for timbre perception in electric hearing," J. Speech Lang. Hear. Res. **54**, 981–994.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). "Caractérisation du timbre des sons complexes. II. Analysis acoustiques et quantification psychophysique (Characterizing the timbre of complex sounds. II. Acoustic analyses and psychophysical quantification)," J. Phys. (Paris) **4**, 625–628.

Krumhansl, C. L. (**2000**). "Rhythm and pitch in music cognition," Psychol. Bull. **126**, 337–340.

Limb, C. J. (**2006**). "Cochlear implant-mediated perception of music," Curr. Opin. Otolaryngol. Head Neck Surg. **14**, 337–340.

Marozeau, J., and De Cheveigné, A. (**2007**). "The effect of fundamental frequency on the brightness dimensions of timbre," J. Acoust. Soc. Am. **121**, 383–387.

Marozeau, J., De Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

Martin, K. D. (**1999**). "Sound-source recognition: A theory and computational model," Ph.D. Dissertation, Massachusetts Institute of Technology, Massachusetts, pp. 1–172.

McAdams, S., Beauchamp, J. W., and Meneguzzi, S. (**1999**). "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," J. Acoust. Soc. Am. **105**, 882–897.

McAdams, S., Winsberg, S., Donaadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of the synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.

Oberfeld, D., Heeren, W., Rennies, J., and Verhey, J. (**2012**). "Spectro-temporal weighting of loudness," PLoS ONE **7**, e50184.

Peretz, I., and Coltheart, M. (**2003**). "Modularity of music processing," Nat. Neurosci. **6**, 688–691.

Peretz, I., and Hyde, K. L. (**2003**). "What is specific to music processing? Insights from congenital amusia," Trends Cogn. Sci. **7**, 362–367.

Russo, F. A., and Thompson, W. F. (**2005**). "An interval size illusion: The influence of timbre on the perceived size of melodic intervals," Percept. Psychophys. **67**, 559–568.

Scharf, B. (**1962**). "Loudness summation and spectrum shape," J. Acoust. Soc. Am. **34**, 228–233.

Shower, E. G., and Biddulph, R. (**1931**). "Differential pitch sensitivity of the ear," J. Acoust. Soc. Am. **3**, 275–287.

Singh, P. G., and Hirsh, I. J. (**1992**). "Influence of spectral locus and F0 changes on the pitch and timbre of complex tones," J. Acoust. Soc. Am. **92**, 2650–2661.

Verhey, J. L., and Kollmeier, B. (**2002**). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am. **111**, 1349–1358.

Vurma, A., and Ross, J. (**2007**). "Timbre-induced pitch deviations of musical sounds," J. Interdiscipl. Music Stud. **1**, 33–50.

Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics—Facts and Models*, 2nd ed. (Springer-Verlag, Berlin), pp. 204–238.