Journal of Experimental Psychology:
Human Perception and Performance
2010, Vol. 36, No. 6, 1631–1644

0096-1523/10/$12.00 DOI: 10.1037/a0018389

Loudness Change in Response to Dynamic Acoustic Intensity

Kirk N. Olsen, Catherine J. Stevens, and Julien Tardieu
University of Western Sydney

Three experiments investigate psychological, methodological, and domain-specific characteristics of loudness change in response to sounds that continuously increase in intensity (up-ramps), relative to sounds that decrease (down-ramps). Timbre (vowel, violin), layer (monotone, chord), and duration (1.8 s, 3.6 s) were manipulated in Experiment 1. Participants judged global loudness change between pairs of spectrally identical up-ramps and down-ramps. It was hypothesized that loudness change is overestimated in up-ramps, relative to down-ramps, using simple speech and musical stimuli. The hypothesis was supported and the proportion of up-ramp overestimation increased with stimulus duration. Experiment 2 investigated recency and a bias for end-levels by presenting paired dynamic stimuli with equivalent end-levels and steady-state controls. Experiment 3 used single stimulus presentations, removing artifacts associated with paired stimuli. Perceptual overestimation of loudness change is influenced by (1) intensity region of the dynamic stimulus; (2) differences in stimulus end-level; (3) order in which paired items are presented; and (4) duration of each item. When methodological artifacts are controlled, overestimation of loudness change in response to up-ramps remains. The relative influence of cognitive and sensory mechanisms is discussed.

Keywords: acoustic intensity, loudness change, dynamics, music perception, auditory looming

Acoustic intensity is the dimension of sound most related to the psychological attribute of loudness (Moore, 2003). The *dynamic* characteristics of loudness perception are fundamental in real-world listening contexts. For example, Huron's (1991, 1992) discussion of the "ramp archetype" exemplifies the use of dynamics in Western tonal music: a predominance of gradual and extended crescendos that maintain the listener's attention, contrasting with short, abrupt diminuendos. Dynamic acoustic structures in musical composition and performance result in, among other things, acoustic intensity change that is likely to correlate perceptually with subjective loudness change. The relationship between acoustic intensity dynamics and loudness is an important issue that has played a secondary role to studies that model loudness using steady-state sounds (e.g., Moore & Glasberg, 2004; Moore, Glasberg, & Baer, 1997; Zwicker & Scharf, 1965), although these sophisticated computational models have begun to incorporate temporal variability inherent in dynamic stimuli (Chalupper & Fastl, 2002; Glasberg & Moore, 2002; Grimm, Hohmann, & Verhey, 2002; Zhang & Zeng, 1997; Zwicker, 1977). Nevertheless, stimuli from real-world listening domains such as speech and music are scarcely investigated in loudness research, especially under dynamic stimulus conditions (but see Neuhoff, 1998, 2001; Skovenborg & Nielsen, 2004).

This article was published Online First September 6, 2010.

Kirk N. Olsen and Julien Tardieu, MARCS Auditory Laboratories, University of Western Sydney; and Catherine J. Stevens, MARCS Auditory Laboratories and School of Psychology, University of Western Sydney.

This work was supported by an Australian Postgraduate Award held by the first author and an Australian Research Council Discovery Project grant (DP0771890) held by the second author. We thank Emery Schubert, Michael Hall, Robert Teghtsoonian, two anonymous reviewers, Roger Dean, and members of the MARCS Music, Sound, and Action Group for helpful comments on an earlier draft. Thanks to John Neuhoff for advice on stimulus generation and programming.

Correspondence concerning this article should be addressed to Kirk N. Olsen, MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, South Penrith, NSW 1797, Australia. E-mail: k.olsen@uws.edu.au

Perception of Dynamic Acoustic Intensity

A sound's dynamic range of intensity can be varied to produce an up-ramp or down-ramp structure. For the purpose of the present paper, an up-ramp refers to a sound that increases continuously in level, whereas a down-ramp decreases continuously in level. Both up-ramps and down-ramps have been constructed using a variety of carrier modulations, such as linear, raised cosine, and inverse-square functions, and can be presented in isolation from one another (i.e., each trial represents one stimulus item with one direction of intensity change, termed hereafter a *single stimulus paradigm*) or as paired items for relative comparisons (termed hereafter a *paired stimulus paradigm*). Although up-ramps and down-ramps are equivalent to each other with respect to long-term energy spectra, stimulus duration, range, and region of intensity sweep, they are perceived differently from each other when participants judge timbre (Irino & Patterson, 1996; Patterson, 1994a, 1994b), duration (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2006; Ries, Schlauch, & DiGiovanni, 2008; Schlauch, Ries, & DiGiovanni, 2001), overall loudness (Ries et al., 2008; Stecker & Hafter, 2000; Susini, McAdams, & Smith, 2007), and loudness change (Neuhoff, 1998, 2001; Seifritz et al., 2002). One issue that dominates these studies is that up-ramps hold greater perceptual salience than down-ramps. In the context of loudness change, this conjecture is surrounded by ongoing conceptual discrepancies that will now be reviewed, so that strengths and weaknesses can be understood and important mechanisms identified.

There is a growing number of studies reporting an overestimation in global judgments of loudness change (i.e., a single, direct post-stimulus judgment of change) for up-ramps, relative to down-ramps. For example, Neuhoff (2001) investigated loudness change using a dynamic vowel stimulus (which sounds like the "a" in "about") and white noise in a paired-stimulus paradigm. Specifically, adults were presented with linear up-ramp/down-ramp and down-ramp/up-ramp pairs, with each item in a pair lasting 1.8 s and separated by a silent interstimulus interval of 0.5 s. Each item covered a 30-dB sound pressure level (SPL) sweep range; for example, 40−70/70−40-dB SPL and 60−90/90−60-dB SPL. Up-ramp vowel stimuli were perceived to change more in loudness than down-ramp stimuli.

In conjunction with equivalent results in a study using pure tones and a vowel timbre (Neuhoff, 1998), Neuhoff claimed that the overestimation of loudness change for up-ramps, relative to down-ramps, is evidence of a "perceptual bias for rising intensities." Neuhoff's evolutionary hypothesis states that a perceptual bias may provide a selective advantage for an organism able to specify the direction or location of a looming (i.e., approaching) source in the environment and underestimate its time-to-contact; that is, perceived contact is earlier than actual contact to allow extra time for appropriate response. The overestimation of loudness change for up-ramps was not of the same magnitude in conditions of white noise because, according to Neuhoff, white noise is rarely found in an ecological setting and may be interpreted as a multiple sound source (e.g., crowd noise) (see also, Ghazanfar, Neuhoff, & Logothetis, 2002).

An alternative view from studies using dynamic stimuli is that the overestimation of loudness change for tones of increasing intensity is due to other biases or artifacts that arise from the method of stimulus presentation. For example, Teghtsoonian, Teghtsoonian, and Canévet (2005) challenge Neuhoff's (1998, 2001) evolutionary hypothesis. In a single-stimulus paradigm and measuring global judgments of loudness change using loudness magnitude estimation (Stevens, 1956), it was demonstrated that as the intensity of an up-ramp's offset increases (60-dB SPL to 90-dB SPL), an "end-level" effect occurs and the overestimation of loudness change in response to up-ramps significantly increases. Thus, the perceptual bias for rising intensities (Neuhoff, 2001) may be an artifactual response mediated by a bias for end-levels (Teghtsoonian et al., 2005). In a paired-stimulus paradigm, the relative influence of intensity region and its consequences for an end-level bias has yet to be investigated with perceived loudness change. The present study will empirically address this conceptual issue.

A related but independent explanation based on end-level differences is one that originally explained global judgments of loudness, and not loudness change per se (Susini et al., 2007). However, it directly applies to the present discussion of loudness change. Susini et al. (2007) argue that a recency effect (Jahnke, 1963) dominated by the latter portion of a sound sequence can explain why up-ramps are louder than down-ramps. When participants are asked to make overall judgments of loudness for a 60−80-dB SPL up-ramp and an 80−60-dB SPL down-ramp, an overestimation of loudness for up-ramps reflects a judgment based on the most recent portion of the sound. The up-ramp finishes on a level 20 dB greater than the down-ramp, and thus will be perceived as louder.

The hypotheses of recency in memory (Susini et al., 2007) and a bias for end-levels (Teghtsoonian et al., 2005) have been proposed in the context of single-stimulus paradigms. Both suggest that the direction of intensity change in itself is not the most salient aspect of the perceptual bias for rising intensities assumed by Neuhoff (2001). A recency effect that leads to the overestimation of loudness change for the second item in a paired-stimulus paradigm may interact with an end-level bias and so provide an alternative explanation of Neuhoff's (2001) apparently robust bias for rising intensities. The contribution of a paired-stimulus paradigm to cognitive and methodological biases will be addressed in the present study.

As the primary focus here is to investigate intensity and loudness from a dynamic standpoint using simple speech and musical stimuli, the present study builds on Neuhoff's (2001) paradigm and addresses the issues of design that underpin the conceptual differences previously discussed. This will be accomplished by (1) using a variety of intensity sweep regions; (2) introducing up-ramps and down-ramps with equivalent end-levels; and (3) presenting steady-state controls. By doing so, we are in a better position to: (a) evaluate the alternative arguments put forward by Teghtsoonian et al. (2005) and Susini et al. (2007) in the investigation of a perceptual bias in loudness change; and (b) investigate whether the perceptual bias is domain specific to vowel sounds (Neuhoff, 2001) and pure tones (Neuhoff, 1998), or more general.

Intensity Dynamics and Loudness Change in Music Perception

As an area of auditory perception that is communicative and affective, music offers a domain in which to generalize these phenomena to other "real-world" listening contexts. Research investigating cross-cultural music (e.g., Western, Japanese, and Hindustani) has shown that in addition to other psychophysical dimensions such as tempo and timbral complexity, acoustic intensity acts as a reliable cue for the interpretation of music's intended emotion(s) that transcends cultural boundaries (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004). Although a link has been made between dynamic intensity and emotion using affective visual stimuli (Tajadura-Jiménez, Väljamäe, & Vastfjall, 2008), most experimental studies investigating intensity perception using musical stimuli have treated acoustic intensity as a static dimension (e.g., Ilie & Thompson, 2006). Considering intensity and loudness in this manner denies the dynamic properties of music that change through time. Thus, the use of intensity dynamics in a simple musical context offers a small, but important step in research investigating the perceptual overestimation of loudness change in this domain.

Speech and musical timbres. In a natural environmental context, sustained sounds that resemble the pure tone and synthetic vowel stimuli used by Neuhoff (1998, 2001) are mainly found in animal vocalizations (Bregman, 1999). These sounds—which are more steady and coherent than white noise—may help segregate a single source from the auditory scene, whereas white noise may indicate sounds from various positions and lack the specificity necessary to direct source perception, location and subsequent action. The comparable effects of pure tones and vowel sounds on judgments of loudness change reported by Neuhoff (1998, 2001) may therefore stem from their spectral structure: vowels are continuous and temporally steady vocalizations (Handel, 1989). As vowels are found in languages across many cultures (Ladefoged & Maddieson, 1996), they are an ecologically comparable substitute to the pure, steady tones rarely found in the natural environment.

One way to increase stimulus complexity and retain ecological validity is to use a simple musical stimulus, such as a violin timbre. The violin is characterized by various vibration modes from the excitation of its strings, creating tones that are stable over time—much like a vowel and pure tone (Handel, 1989; Rossing, Moore, & Wheeler, 2002). Furthermore, as a source of sound production, the voice and violin share key properties (Askenfelt, 1991). From a source/filter perspective (Fant, 1970), Askenfelt (1991) argues that the sound source of a stringed instrument (the bow/string coupling) and the voice (larynx) are both separated from the filtering mechanism; in this case, the sound box (i.e., the body of a violin) and the vocal tract, respectively. Among other things, this source/filter separation enables the "partials to enter and leave the resonance peaks in the filter according to the selected fundamental frequency" (Askenfelt, 1991, p. 254). Secondly, a sound's spectral balance from the voice and violin can be independently controlled from the dynamic level. Although the interaction between spectral balance and dynamic level is beyond the scope of the present study, the similarities in sound production systems across speech and stringed instruments suggests that an overestimation of loudness change in response to up-ramp stimuli will be evident when a violin timbre is presented.

Monotone and chord structures. Manipulating the texture (i.e., number of layers) of vowel and violin stimuli is a small, yet important step for development into the realm of music. Both vowel and violin stimuli can be investigated using the multilayered quality of a musical chord. Although investigations of textural qualities and loudness have received less empirical attention than other acoustic parameters in music perception, an increase in stimulus texture within musical excerpts has been linked with self-reported arousal (Kellaris & Kent, 1993; Schubert, 2004) and could plausibly interact with intensity change to enhance arousal and associated experience. Anticipation of the effect of more complex layering of sound can be illuminated with consideration of psychological tension and musical expectation (Huron, 2006).

In a musical context, one way of building tension to a heightened point of expectation involves an increase in the overall "size" of the sound (Cabrera, 1999); specifically, where a dynamic increase in intensity is coupled with the augmentation of sound sources, creating an additive effect (Huron, 2006). Furthermore, an event of increasingly heightened tension in tonal music is often characterized by the use of subjectively "harsh" sounds such as dissonant tones or chords. Therefore, a musical climax that follows an increased build up of tension is often spectrally associated with dissonant sounds and a dynamic increase of acoustic intensity. This scenario will be operationalized by using a multilayered sound structure of the dissonant diminished triad chord (C, E^b, G^b) that rises in intensity, the outcome of which may produce a microcosm of a musically climactic event, whereby characteristic perceptual and physiological response patterns of heightened tension should be evident and more pronounced from multilayered stimuli, relative to a monotone version of the same timbre. In the current perceptual study, we anticipate a greater overestimation of loudness change for up-ramp chords, relative to up-ramp monotone stimuli.

Stimulus duration. Tacit in the 1.8-s up-ramp duration that forms the empirical basis for Neuhoff's (1998, 2001) evolutionary hypothesis is the rate of intensity change. A 1.8-s sound that rises over a 30-dB SPL sweep size changes at a specific rate for those parameters. However, a sound that rises over the same sweep size, but over a longer duration will manifest as a slower rate of change. For example, a linear change of 30-dB SPL over 1.8-s changes at a rate of 16.67-dB SPL per second, whereas a linear change of 30-dB SPL over 3.6-s changes at a rate of 8.33-dB SPL per second. Schubert and Dunsmuir's (1999) work on the continuous measurement of emotional arousal and qualities of loudness has shown that the more sudden a change in loudness (e.g., $1-2$ s compared to $2-3$ s), the faster the change in reported emotional arousal. In vocal production, sound sequences that become faster are more likely to elicit high arousal (Scherer & Oshinsky, 1977). If arousal increase is correlated with perception and rate of intensity change, then differences in judgments of loudness change for pairs of 1.8-s up-ramps and down-ramps will be greater than those for 3.6-s pairs. Therefore, it is hypothesized that, because of a faster rate of intensity change, 1.8-s stimulus presentations elicit a greater overestimation of loudness change for up-ramps, relative to 3.6-s up-ramp presentations.

The aim of Experiment 1 was to investigate global judgments of loudness change in response to continuous intensity change within a simple and controlled musical context. Experiment 2 was designed to further investigate global judgments of loudness change by introducing additional control conditions and comparisons; that is, controlling for end-level differences between up-ramps and down-ramps, while elucidating further the relative contribution of recency in a paired-stimulus paradigm. Experiment 3 used a single-stimulus paradigm to investigate and clarify further the significant results and paired-stimulus artifacts from Experiment 2. Specifically, we ask, is it the case that in some instances of music perception a continuous increase in intensity is overestimated in loudness change? Does this perceptual overestimation hold when the possibility of methodological biases and response constraints are controlled?

Experiment 1: Musical Timbre, Chords, and Loudness Change

Experiment 1 was designed to investigate previous work reporting a perceptual bias for rising intensities (Neuhoff, 2001) using simple speech and musical timbres. It was realized as a $2 \times 2 \times 2$ within-subjects factorial design. The independent variables consisted of timbre (vowel and violin), layer (monotone and chord), and duration of stimulus presentation (1.8 s and 3.6 s). The dependent variable followed Neuhoff (2001), where relative judgments of loudness change between pairs of items were made in each trial. In a paired-stimulus paradigm such as this, either item is judged to change more in loudness than the other, or both items are judged to change the same amount in loudness. When an increasing intensity item (up-ramp) is perceived to change more in loudness than a down-ramp, an increasing response is recorded. When a decreasing intensity item (down-ramp) is perceived to change more in loudness than an up-ramp, a decreasing response is recorded. A no-difference response is recorded if there is no difference in loudness change perceived between the two dynamic items.

It was hypothesized that (1) a greater proportion of up-ramps are judged to change more in loudness than down-ramps when presented in vowel monotone 1.8-s and violin monotone 1.8-s conditions; (2) a greater proportion of up-ramps are judged to change more in loudness in the vowel 1.8-s and violin 1.8-s dissonant diminished triad chord (C, E^b, G^b) conditions, relative to up-ramps in their 1.8-s monotone counterparts; and (3) a greater proportion of up-ramps are judged to change more in loudness in 1.8-s stimulus conditions (16.67-dB SPL per second rate of intensity change) relative to up-ramps in the 3.6-s stimulus conditions (8.33-dB SPL per second rate of intensity change).

Method

Participants

The sample consisted of 32 adult participants recruited from the University of Western Sydney (25 females and 7 males; M = 20.06 years, SD = 2.33, range = 17−25 years). All reported normal hearing. Seven participants had received minimal individual musical training (M = 1.29 years, SD = .57, range = .5−2 years).

Stimuli and Equipment

All dynamic stimuli followed a linear intensity increase (up-ramp) or decrease (down-ramp) from 60−90-dB SPL and 90−60-dB SPL, respectively. The generation of vowel stimuli began with a 1.8-s and a 3.6-s steady-state synthetic vowel (/ə/) from a Klatt synthesizer[1] (Klatt, 1980) using the default sampling frequency of 8 kHz. A recorded violin sound (default sampling frequency of 44.1 kHz) from a LogicPro (Version 7.2.3) EXS24 integrated sampler was used for the steady-state violin stimuli over 1.8-s and 3.6-s durations. Up-ramps and down-ramps were constructed from the steady-state exemplars in a sound-attenuated booth using a custom computer program written in MAX-MSP (Version 4.6.3). With the aid of an Ono Sokki LA-1210 Sound Level Meter microphone placed 13 mm from the centre of the headphone speaker element, a minimum (60 dB) and maximum (90 dB) intensity level was recorded in the MAX-MSP program using each steady-state sound. The program generated an up-ramp and a down-ramp for each condition by using the two recorded dB levels as onset/offset anchors and creating a linear change between them using each original steady-state sound. The new dynamic stimuli were imported into Audacity (Version 1.3.3) sound editing program and a 40-ms fade-in and fade-out was incorporated to remove any onset/offset clicks. Single sweep stimuli were spliced together to form up-ramp/down-ramp and down-ramp/up-ramp orders of each condition. A 0.5-s silent interstimulus interval separated each item in a pair. All vowel stimuli were characterized with the C_3 fundamental frequency to correspond closely to the vowel stimuli used by Neuhoff (2001). As the violin's frequency range does not extend to $C_{3,}$ the C_4 fundamental frequency was used. Accordingly, the chord structures of the dissonant diminished triad for the vowel was constructed from C_3 (F_0 = 130.81 Hz), E_3^b (F_0 = 155.56 Hz), and G_3^b (F_0 = 185 Hz); for violin, C_4 (F_0 = 261.63 Hz), E_4^b (F_0 = 311.13 Hz), and G_4^b (F_0 = 369.99 Hz).

The generation and presentation of the computer-based visual analogue scale (VAS) response system, sound randomization, and protocol of the experiment was completed with the Music Experiment Development System (Kendall, 2000). Stimuli were presented binaurally through Sennheiser HD 25 headphones. The experiment was conducted in a sound attenuated booth.

Procedure

Participants first read an experiment information sheet, gave written informed consent and received standardized instructions regarding the task. Following the procedure used by Neuhoff (2001), participants were presented with pairs of up-ramp/down-ramp or down-ramp/up-ramp items, counterbalanced to distribute serial order effects, and were not informed of the expected distribution of stimuli. The main task was to indicate whether the amount of loudness change for each item in each trial was the same, or whether one sound changed more in loudness than the other. This response was to be made in a time-frame up to 3 s using the VAS. Participants used a computer mouse to slide a cursor to one of two ends on the VAS, marked as "Sound 1 Changed More" at the far left and "Sound 2 Changed More" at the far right of the bipolar scale. The cursor was not moved if the two sounds were deemed to change equally in loudness and was recorded as a no difference response. The experiment consisted of six practice stimuli, followed by four blocks of 16 randomized experimental trials (total of 64). Overall, each condition was presented eight times. The experiment took approximately 20 minutes.

Results

All statistical comparisons were within-subjects planned contrasts (α = .05) with partial eta squared (η_p^2) as a measure of effect size (Cohen, 1973). Proportional response rates were calculated for each participant by dividing the number of each specific response (i.e., increasing, decreasing, and no difference) by the total number of trials in each condition. The mean proportions for each condition are shown in Table 1.

First it was hypothesized that the vowel monotone 1.8-s and violin monotone 1.8-s up-ramp conditions would be judged to have changed more in loudness than the vowel monotone 1.8-s and violin monotone 1.8-s down-ramp conditions. The proportion of up-ramp stimuli judged to have changed more in loudness (an increasing response) was significantly greater than the proportion of down-ramp stimuli judged to have changed more in loudness (a decreasing response) in both the vowel monotone 1.8-s condition, $F(1, 31) = 114.76$, $p < .001$, $\eta_p^2 = .79$, and the violin monotone 1.8-s condition, $F(1, 31) = 72.52$, $p < .001$, $\eta_p^2 = .70$.

Second, it was hypothesized that when a difference in loudness change is perceived between up-ramps and down-ramps, overestimation in loudness change for up-ramps is greater in the vowel chord 1.8-s and violin chord 1.8-s conditions, relative to up-ramps in the vowel monotone 1.8-s and violin monotone 1.8-s conditions. These hypotheses were investigated by comparing increasing responses between vowel chord 1.8-s and vowel monotone 1.8-s

[1] See < http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html > for the Klatt vowel synthesis interface.

Table 1
Experiment 1: Mean Increasing, Decreasing, and No Difference Responses Shown as Proportions as a Function of Condition

Condition	Increasing	Decreasing	No difference
Vowel 1.8 s			
Monotone	.73 (.20)	.14 (.16)	.13 (.18)
Chord	.71 (.20)	.12 (.15)	.18 (.19)
Vowel 3.6 s			
Monotone	.79 (.18)	.10 (.15)	.11 (.15)
Chord	.77 (.17)	.13 (.14)	.11 (.14)
Violin 1.8 s			
Monotone	.61 (.24)	.13 (.15)	.26 (.23)
Chord	.68 (.21)	.16 (.18)	.16 (.18)
Violin 3.6 s			
Monotone	.66 (.16)	.15 (.15)	.19 (.19)
Chord	.72 (.18)	.16 (.14)	.13 (.17)

Note. Standard deviations are shown in parentheses; not all totals sum to 1 due to decimal rounding.

conditions, and between violin chord 1.8-s and violin monotone 1.8-s conditions. As can be seen in Table 1, there was no significant difference between vowel chord 1.8-s and vowel monotone 1.8-s conditions, $F(1, 31) = .50$, $p > .05$, $\eta_p^2 = .02$, or between violin chord 1.8-s and violin monotone 1.8-s conditions, $F(1, 31) = 1.87$, $p > .05$, $\eta_p^2 = .06$.

Finally, it was hypothesized that when a difference in loudness change is perceived between up-ramps and down-ramps, overestimation in loudness change for up-ramps is greater in the 1.8-s conditions, relative to up-ramps in the 3.6-s conditions. This was examined by comparing increasing responses between 1.8-s and 3.6-s conditions. The mean proportion of increasing responses in the 1.8-s conditions was .68 ($SD = .16$), significantly less than the proportion of increasing responses in the 3.6-s conditions ($M = .73$; $SD = .13$), $F(1, 31) = 5.17$, $p < .05$, $\eta_p^2 = .14$, thus providing a result in the direction opposite to the hypothesis.

Post-Hoc Stimulus Order Effects

In a paired-stimulus paradigm, the order of stimulus presentation has been found to significantly affect loudness (e.g., Stecker & Hafter, 2000). Stecker and Hafter report that 250-ms pure tone up-ramps are louder than 250-ms pure tone down-ramps, only when the up-ramp is presented as the second item in a paired-stimulus sequence. Figure 1 shows the proportion of increasing, decreasing, and no-difference responses for up-ramp/down-ramp and down-ramp/up-ramp orders of stimuli in the vowel monotone 1.8-s and violin monotone 1.8-s conditions. For responses overestimating loudness change for up-ramps (an increasing response), there is a significant difference between the two orders of paired-stimuli in the vowel monotone 1.8-s condition, $F(1, 31) = 10.09$, $p < .01$, $\eta_p^2 = .25$, and the violin monotone 1.8-s condition, $F(1, 31) = 16.24$, $p < .001$, $\eta_p^2 = .34$. However, in opposition to results of Stecker and Hafter's (2000), a significant overestimation of loudness change for up-ramps, relative to down-ramps, still holds when a down-ramp follows an up-ramp in the vowel monotone 1.8-s condition, $F(1, 31) = 25.75$, $p < .001$, $\eta_p^2 = .45$, and the violin monotone 1.8-s condition, $F(1, 31) = 21.46$, $p < .001$, $\eta_p^2 = .41$.

Discussion

Experiment 1 investigated the role of intensity dynamics in listeners' judgments of loudness change for simple musical and nonmusical stimuli. It was hypothesized that a greater proportion of up-ramps are judged to change more in loudness than down-ramps, when presented in vowel monotone 1.8-s and violin monotone 1.8-s conditions. This hypothesis was supported in both the vowel and violin conditions. The overestimation of loudness change for up-ramp stimuli reported by Neuhoff (2001) was replicated using a synthetic vowel timbre and recovered with comparable results using a violin timbre.

Secondly, it was hypothesized that the vowel 1.8-s and violin 1.8-s timbres of a dissonant diminished triad chord (C, E^b, G^b) would elicit a greater overestimation in loudness change for up-ramps, relative to monotone stimulus up-ramps. Results indicate that the addition of stimulus layers for vowel and violin stimuli did not enhance loudness change overestimation of up-ramps. Therefore, chord up-ramps were overestimated in loudness change, relative to chord down-ramps, but the degree of overestimation

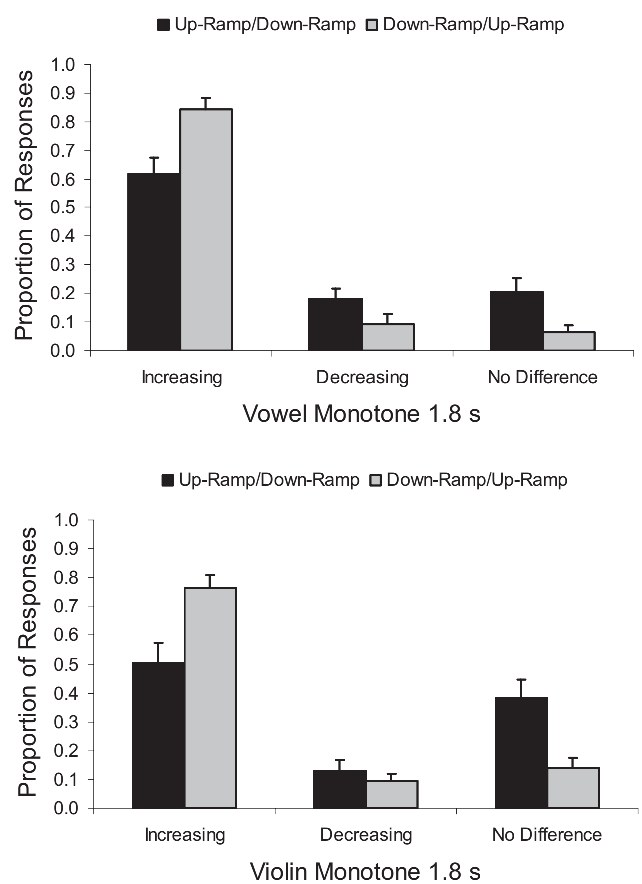

Figure 1. Experiment 1 mean proportion of increasing, decreasing, and no difference responses indicating that an increasing (up-ramp) or decreasing (down-ramp) intensity stimulus changed more in loudness than the other, or alternatively, that there was no difference in judged loudness change. Upper panel shows vowel and lower panel shows violin monotone 1.8-s conditions, with sweep sizes of 30-dB SPL. Error bars represent standard error of the mean.

was not greater than that of monotone stimuli. Future studies could investigate the effect of the diminished triad chord further by contrasting it with a major triad chord and thus control the bandwidth of each stimulus.

The final hypothesis stated that the 1.8-s stimulus duration would elicit a greater overestimation in loudness change for up-ramps, relative to the longer 3.6-s up-ramps, because of a faster rate of intensity change (16.67-dB SPL per second and 8.33-dB SPL per second, respectively). Opposite to this prediction, 3.6-s up-ramp stimuli were overestimated in loudness change at a significantly greater proportion than up-ramps in 1.8-s conditions. Stimulus duration and direction of intensity change may interact to increase loudness change for up-ramps as stimulus duration increases. One explanation here is that the musically untrained participants were unable to separate judgments of loudness change from duration. A similar result of stimulus duration on global judgments of loudness in response to up-ramps and down-ramps has been reported by Susini and colleagues (Susini, McAdams, & Smith, 2002; Susini et al., 2007). Although these authors did not measure loudness change per se, it would seem that duration plays a significant role in any asymmetry in global judgments of dynamic intensity sweeps with durations greater than 2 s. Indeed, these dimensions may be processed as integral, rather than separable (Garner & Felfoldy, 1970; Melara & Marks, 1990).

The stimulus order effect illustrated in Figure 1 shows that the paired-stimulus order significantly affects the overestimation of loudness change for up-ramp vowel monotone 1.8-s and violin monotone 1.8-s stimuli. Listeners' perceptions may be biased toward the latter portion of the second sound in each trial (which finished at 90-dB SPL 50% of the time and 60-dB SPL 50% of the time). Hence, the overestimation of loudness change for up-ramps is accentuated when an up-ramp follows a down-ramp, because the end-level of the up-ramp is 30-dB SPL greater than a down-ramp. A recency effect biasing judgments toward the end-level of each stimulus can explain this effect (Susini et al., 2007).

Experiment 1 was not designed to control for the 30-dB SPL difference in end-levels between the 60−90-dB SPL region. How is the strength of the perceptual overestimation affected when end-levels are balanced between both dynamic items? Furthermore, does the overestimation of loudness change for up-ramps diminish as the intensity region of each sweep decreases? Experiment 2 was designed to address these questions investigating the effect of intensity region and end-level differences on loudness change overestimation.

Experiment 2: Effect of End-Level and Intensity Region

The main aim of Experiment 2 was to investigate the end-level bias reported by Teghtsoonian et al. (2005) and the role of recency in a paired stimulus paradigm. A 2 × 2 × 4 within-subjects factorial design was realized with two levels of timbre (vowel, violin), two levels of duration (1.8 s, 3.6 s) and four levels of intensity sweep region. These sweep regions were a dynamic high (DH) condition (70−90/90−70-dB SPL), a dynamic low (DL) condition (50−70/70−50-dB SPL), a dynamic balanced (DB) condition (50−70/90−70-dB SPL), and a no change (NC) control condition (70−70/70−70-dB SPL). As in Experiment 1, the DV was the mean proportion of increasing, decreasing and no difference responses.

The DL and DH conditions were designed to investigate global judgments of loudness change using two intensity-sweep regions. The difference in loudness change overestimation between the two intensity regions was of specific interest. The end-level argument from Teghtsoonian et al. (2005) states that overestimation of loudness change for up-ramps increases as sweep region and end-level increases. Therefore, overestimation of loudness change for up-ramps was expected to be of a greater proportion in DH conditions, relative to up-ramps in DL conditions.

A difference in loudness change between DL and DH intensity sweep regions can be further discussed in the context of perceptual decruitment (e.g., Canévet & Scharf, 1990; Canévet, Teghtsoonian, & Teghtsoonian, 2003; Teghtsoonian, Teghtsoonian, & Canévet, 2000). From methods such as loudness magnitude estimation in a single-stimulus paradigm, decruitment occurs when participants perceive the change in loudness of a pure tone down-ramp to be greater than loudness change in response to single sound bursts representing onset and offset levels of the down-ramp. The effect is most pronounced when the intensity sweep of a down-ramp nears 40-dB SPL and below. In the paired-stimulus paradigm, the end-level intensity of a down-ramp in DL conditions approaches the critical point for decruitment. Coupled with a paired-stimulus order effect that may be influenced by a recency bias, a decruitment-like effect could manifest in DL conditions as an overestimation of loudness change for down-ramps when presented as the second and thus most recent item in a paired-stimulus sequence. This effect would likely disappear for down-ramps in the corresponding up-ramp/down-ramp sequence in DH conditions, as the 90−70-dB SPL region is too great to elicit decruitment.

In addition, an overestimation in loudness change for up-ramps, relative to down-ramps, and a stimulus order effect between increasing and decreasing responses in DB conditions, would provide counterevidence for recency based on end-level differences (Susini et al., 2007). This is because the end-level of each item is 70-dB SPL. However, support for these predictions does not rule out the possibility that recency could still play a significant role in the perception of the entire stimulus, and not merely end-level differences.

Finally, as shown in Experiment 1, the duration of stimulus presentation significantly increases loudness change overestimation in response to up-ramps. In Experiment 2, the NC conditions were designed to shed light on the effect of stimulus duration on loudness change. With no intensity change in NC conditions, a response judging the first sound to change more in loudness than the second was coded as a "Sound 1" response; a response judging the second sound to change more in loudness than the first was coded as a "Sound 2" response; and if the two identical items were judged to change equally in loudness, a no difference response was coded. If an increase in stimulus duration and thus the duration of each trial biases judgments toward the second sound in each pair, then we would expect that the overestimation of loudness change in response to pairs of steady-state items would be greater in proportion for Sound 2, relative to Sound 1, in the 3.6-s NC conditions, relative to the 1.8-s NC conditions.

Specifically, it was hypothesized that (1) up-ramp items are overestimated in loudness change at a greater proportion than down-ramps in all DL and DH conditions; (2) a greater proportion of up-ramps are judged to change more in loudness in DH conditions, relative to up-ramps in DL conditions; (3) down-ramps are overestimated in loudness change, relative to up-ramps, in DL conditions in the up-ramp/down-ramp paired-stimulus sequence; (4) up-ramps are overestimated in loudness change at a greater proportion than down-ramps in DB conditions; and (5) in NC conditions, Sound 2 items are perceived to change more in loudness than Sound 1 items at a greater proportion for 3.6-s conditions, relative to 1.8-s conditions.

Method

Participants

The sample consisted of 32 adult participants (25 females and 7 males; M = 19.88 years, SD = 2.37, range = 18−25 years) recruited from the University of Western Sydney and who did not participate in Experiment 1. All reported normal hearing. Eleven participants had received minimal individual musical training (M = 1.06 years, SD = .39, range = .5−1.5 years).

Stimuli and Equipment

Stimulus generation was identical to Experiment 1, noting that the NC condition did not require the generation of an up-ramp or down-ramp. As there were no statistical differences between the monotone and chord results in Experiment 1, only monotone stimuli were presented. Equipment was identical to Experiment 1.

Procedure

The procedure closely followed Experiment 1. However, the order of the bipolar anchors on the VAS was reversed for every other participant to distribute any response bias towards a particular end of the scale. Four practice stimuli were first presented to participants, followed by four blocks of 32 randomized experimental trials (total of 128), counterbalanced to distribute serial order effects. In total, eight trials of each condition were presented to each participant. The experiment took approximately 30 minutes.

Results

For all dynamic conditions, Table 2 reports descriptive statistics for the proportion of responses indicating: (1) *increasing* intensity items changed more in loudness than decreasing intensity items; (2) *decreasing* intensity items changed more in loudness than increasing intensity items; or (3) *no difference* in loudness change was perceived between increasing and decreasing stimuli. To address the specific hypotheses, all statistical comparisons were within-subjects contrasts (α = .05) with partial eta squared (η_p^2) as a measure of effect size (Cohen, 1973).

First, it was hypothesized that a significant difference in the overestimation of loudness change for up-ramps, relative to down-ramps, occurs in DL and DH conditions. Figure 2 shows that for the DH condition, a significant difference in loudness change between increasing and decreasing responses was observed across all conditions. Reliably, these results are similar to the monotone conditions in Experiment 1. For DL conditions, a significant difference in loudness change between increasing and decreasing responses was observed only for the vowel 1.8-s condition, $F(1, 31) = 4.96$, $p < .05$, $\eta_p^2 = .14$, and vowel 3.6-s condition, $F(1, 31) = 28.85$, $p < .001$, $\eta_p^2 = .48$. Thus, the hypothesis was partially supported.

Second, it was hypothesized that a greater proportion of up-ramps are judged to change more in loudness in DH conditions, relative to the DL conditions. This hypothesis was examined by comparing increasing responses in DL conditions with increasing responses in DH conditions. As predicted, there was a significantly greater proportion of increasing responses to the 70−90-dB SPL DH intensity region (M = .73; SD = .15), relative to increasing

Table 2
Experiment 2: Mean Increasing, Decreasing, and No Difference Responses Shown as Proportions

	Up-ramp/Down-ramp sequence			Down-ramp/Up-ramp sequence			Combined sweep presentations		
Condition	Increasing	Decreasing	No difference	Increasing	Decreasing	No difference	Increasing	Decreasing	No difference
Vowel 1.8 s									
DL	.28 (.27)	.47 (.33)	.25 (.27)	.67 (.30)	.13 (.21)	.20 (.24)	.47 (.23)	.31 (.22)	.22 (.21)
DH	.55 (.28)	.20 (.24)	.25 (.25)	.84 (.27)	.13 (.21)	.04 (.09)	.70 (.22)	.16 (.16)	.14 (.16)
DB	.38 (.23)	.37 (.24)	.25 (.30)	.46 (.31)	.36 (.29)	.18 (.22)	.42 (.21)	.37 (.22)	.22 (.23)
Vowel 3.6 s									
DL	.30 (.27)	.60 (.27)	.10 (.15)	.91 (.18)	.06 (.14)	.03 (.08)	.60 (.15)	.34 (.14)	.07 (.08)
DH	.66 (.31)	.26 (.27)	.08 (.17)	.96 (.13)	.04 (.13)	.00 (.00)	.81 (.17)	.15 (.16)	.04 (.09)
DB	.48 (.30)	.37 (.32)	.16 (.20)	.66 (.31)	.28 (.31)	.06 (.13)	.56 (.25)	.33 (.26)	.11 (.14)
Violin 1.8 s									
DL	.25 (.21)	.44 (.32)	.31 (.27)	.55 (.29)	.27 (.25)	.19 (.23)	.40 (.18)	.35 (.23)	.25 (.19)
DH	.49 (.31)	.34 (.28)	.17 (.23)	.91 (.19)	.09 (.19)	.01 (.04)	.70 (.17)	.20 (.17)	.09 (.12)
DB	.42 (.25)	.43 (.29)	.15 (.28)	.44 (.28)	.45 (.29)	.12 (.21)	.42 (.19)	.45 (.18)	.13 (.18)
Violin 3.6 s									
DL	.28 (.30)	.63 (.32)	.09 (.19)	.73 (.28)	.20 (.25)	.06 (.13)	.49 (.19)	.42 (.21)	.09 (.12)
DH	.50 (.34)	.38 (.32)	.13 (.14)	.89 (.20)	.09 (.18)	.02 (.07)	.70 (.20)	.23 (.18)	.07 (.08)
DB	.35 (.29)	.48 (.29)	.16 (.21)	.48 (.33)	.45 (.32)	.07 (.16)	.42 (.24)	.46 (.23)	.12 (.16)

Note. Not all totals sum to 1 due to decimal rounding. Standard deviations are shown in parentheses. DL = dynamic low (50–70/70–50-dB SPL); DH = dynamic high (70–90/90–70-dB SPL); DB = dynamic balanced (50–70/90–70-dB SPL).

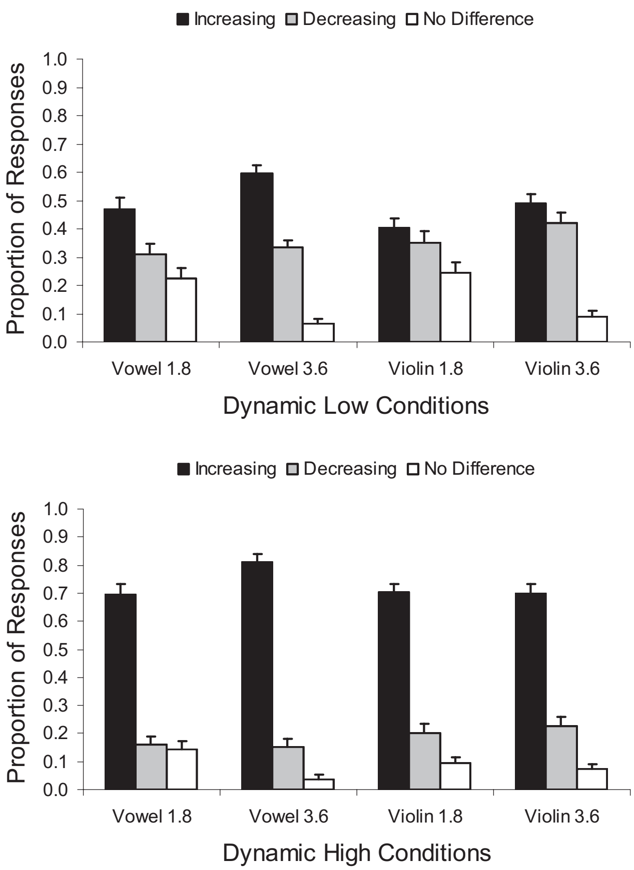

Figure 2. Mean proportion of increasing (up-ramp), decreasing (down-ramp), and no difference responses. Upper panel shows dynamic low (50−70/70−50-dB SPL) and lower panel shows dynamic high (70−90/90−70-dB SPL) conditions for each timbre and duration presentation in Experiment 2. A response of no difference indicated no perceived difference in loudness change between up-ramp and down-ramp sweeps. Error bars represent standard error of the mean.

responses to the 50−70-dB SPL DL intensity region (M = .49; SD = .14), $F(1, 31) = 108.58$, $p < .001$, $\eta_p^2 = .78$. Overestimation of loudness change in response to up-ramps increases as the intensity region of the sweep increases.

Third, it was hypothesized that a greater proportion of down-ramps are overestimated in loudness change (a decreasing response), relative to up-ramps, in the up-ramp/down-ramp sequence of all DL conditions, as decruitment may influence the perception of the down-ramp as intensity falls to 50-dB SPL. This was not expected from the equivalent paired-stimulus sequence in the DH conditions, because of its higher region of intensity change. As can be seen in Figure 3, the difference between increasing and decreasing responses in the DL up-ramp/down-ramp sequence was significant for the violin 1.8-s DL condition, $F(1, 31) = 4.98$, $p < .05$, $\eta_p^2 = .14$, the violin 3.6-s DL condition, $F(1, 31) = 10.81$, $p < .01$, $\eta_p^2 = .26$, and the vowel 3.6-s DL condition, $F(1, 31) = 11.38$, $p < .01$, $\eta_p^2 = .27$. The vowel 1.8-s DL condition approached significance, $F(1, 31) = 3.93$, $p = .056$, $\eta_p^2 = .11$. Thus, in three of four DL conditions, down-ramps were perceived to change more in loudness than up-ramps when they were presented as the second item in a paired-stimulus sequence. This effect is reversed in the equivalent paired-stimulus sequence in DH conditions (Figure 3).

It was further hypothesized that if the overestimation of loudness change is independent of end intensity level, up-ramps are perceived to change more in loudness, relative to down-ramps, in DB conditions. As can be seen in Figure 4, the difference between the increasing and decreasing responses was significantly different in the vowel 3.6-s DB condition only, $F(1, 31) = 7.33$, $p < .05$, $\eta_p^2 = .19$. Partially supporting the hypothesis, a difference in loudness change was observed when end level intensity is balanced between up-ramps and down-ramps. However, post-hoc analysis of this effect in the paired-stimulus sequence (i.e., up-ramp/down-ramp and down-ramp/up-ramp) revealed a significant difference between increasing and decreasing responses in the down-ramp/up-ramp sequence, $F(1, 31) = 12.26$, $p < .001$, $\eta_p^2 = .28$, but not in the up-ramp/down-ramp sequence, $F(1, 31) = 1.09$, $p > .05$, $\eta_p^2 = .01$ (see Figure 5).

Finally, it was hypothesized that when no change in intensity is physically present between two items in a pair (i.e., 70−70/70−70-dB SPL), Sound 2 is perceived to change more in loudness than Sound 1 as stimulus duration increased to 3.6 s, relative to 1.8 s, in vowel and violin conditions. This hypothesis was supported. As can be seen in Figure 6, there was no significant

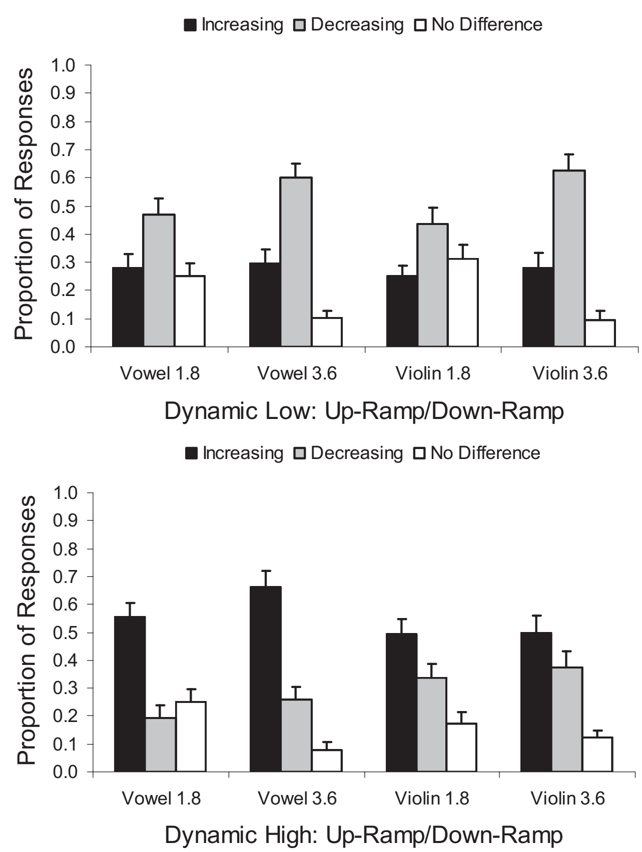

Figure 3. Mean proportion of increasing (up-ramp), decreasing (down-ramp), and no difference responses in the up-ramp/down-ramp paired stimulus sequence for dynamic low (50−70/70−50-dB SPL; upper panel) and dynamic high (70−90/90−70-dB SPL; lower panel) conditions in Experiment 2. Error bars represent standard error of the mean.

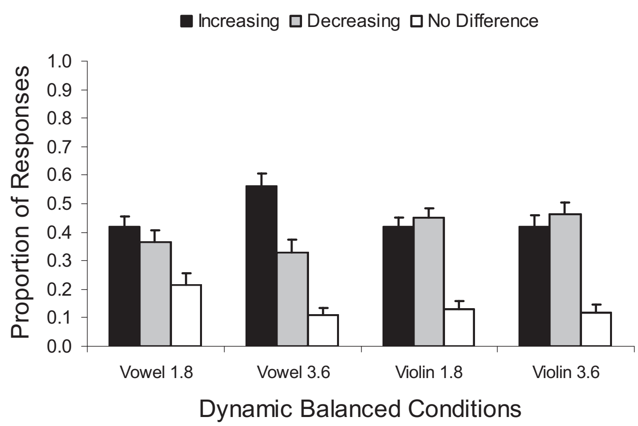

Figure 4. Mean proportion of increasing (up-ramp), decreasing (down-ramp), and no difference responses in the dynamic balanced conditions (50−70-dB SPL up-ramp and 90−70-dB SPL down-ramp) for each timbre and duration presentation in Experiment 2. A response of no difference indicated no difference in judged loudness change between up-ramp and down-ramp sweeps. Error bars represent standard error of the mean.

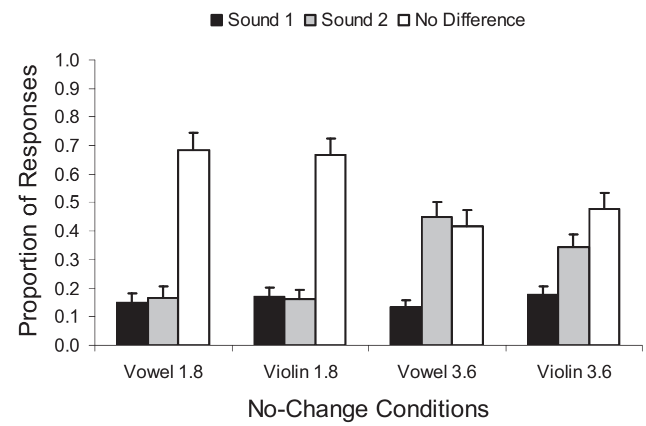

Figure 6. Mean proportion of trials that participants judged the first steady-state item (Sound 1) or the second steady-state item (Sound 2) to have changed more in loudness in the no-change conditions in Experiment 2. A response of no difference indicated no perceived difference in judged loudness change between each item in a pair. Error bars represent standard error of the mean.

difference between Sound 1 and Sound 2 responses in the vowel 1.8-s condition, $F(1, 31) = .22$, $p > .05$, $\eta_p^2 = .01$, and the violin 1.8-s condition, $F(1, 31) = .11$, $p > .05$, $\eta_p^2 = .00$. However, as the duration of each item doubled to 3.6 s, the second item in each pair was perceived to change significantly more in loudness than the first in the vowel 3.6-s condition, $F(1, 31) = 29.57$, $p < .001$, $\eta_p^2 = .49$, and the violin 3.6-s condition, $F(1, 31) = 10.89$, $p < .01$, $\eta_p^2 = .26$.

Discussion

Experiment 2 addressed loudness change in response to dynamic stimuli. First, it was hypothesized that up-ramps are overestimated in loudness change, relative to down-ramps, in high- (DH) and medium-intensity (DL) sweep regions. This hypothesis was partially supported: a significant difference between increasing and decreasing responses was observed in all DH conditions, whereas for the DL conditions, overestimation of loudness change from up-ramps was eliminated with a violin timbre over 1.8-s and 3.6-s durations. The effect of timbre is difficult to explain. A fine-grained acoustic anomaly in the violin stimulus (e.g., a short but noticeable periodic tremolo) may have provided an auditory cue for listeners, regardless of the dynamic intensity characteristics.

Furthermore, as hypothesized, the overestimation of loudness change for up-ramps was of a greater proportion in DH conditions, relative to DL conditions. This result supports the end-level argument proposed by Teghtsoonian et al. (2005). That is, the overestimation of direct judgments of loudness change increases as a function of intensity region: as the end-level of up-ramps increase, so does the overestimation of loudness change. The effect of intensity region on loudness change can also be applied to the evolutionary stance of Neuhoff (1998, 2001) if we consider that, all things being equal, louder sounds are closer and can pose a greater threat. Low-intensity sounds elicit a smaller difference because, in terms of distance, low-intensity sounds would be further away, less salient, and less threatening.

A more striking observation using this paradigm is that down-ramps are overestimated in loudness change, relative to up-ramps, when they are presented as the second item that ends on 50-dB SPL in a paired-stimulus sequence. The mechanism of adaptation that underpins decruitment (Canévet & Scharf, 1990, but see Schlauch, 1992) is likely to have some degree of influence on loudness change in DL conditions. Loudness adaptation is generally represented as a decrease in loudness of a steady-state sound, which becomes more pronounced when stimulus intensity approaches the threshold of human hearing (Scharf, 1983). As a dynamic stimulus, the down-ramps in DL conditions fall to the medium-intensity level of 50-dB SPL and are likely to elicit the adaptation mechanism associated with decruitment.

However, recency based on end-level differences cannot be completely ruled out. The balanced end-level conditions were

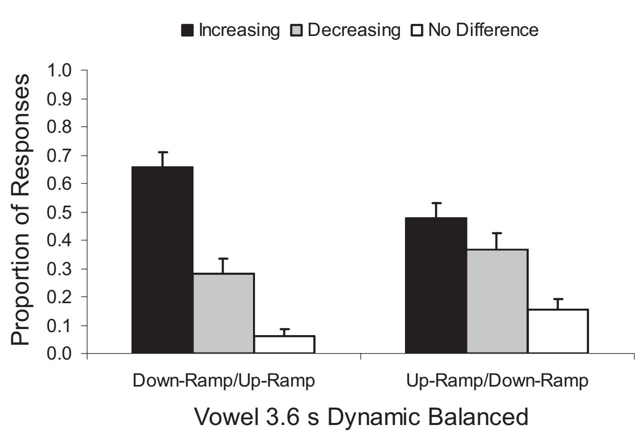

Figure 5. Stimulus order effects for the vowel 3.6-s dynamic balanced condition in Experiment 2. The proportion of increasing, decreasing, and no difference responses are plotted as two sequences of sweep order. Error bars represent standard error of the mean.

designed to explicitly investigate an end-level recency effect by presenting dynamic stimuli that ended on 70-dB SPL. Up-ramps were significantly overestimated in loudness change, relative to down-ramps, in the vowel 3.6-s balanced end-level condition. Judgments of loudness change that are biased by the recency of the end-level of each item cannot explain this significant difference. However, post-hoc analysis of this result revealed that a significant difference in the overestimation of loudness change for up-ramps under balanced end-level conditions is explained by a significant difference in the down-ramp/up-ramp sequence. The fact that judgments of loudness change are biased toward the second of two steady-state 3.6-s items in the NC conditions may help explain why the vowel 3.6-s condition and not the vowel 1.8-s condition elicited an overestimation of loudness change in balanced end-level conditions.

From Experiments 1 and 2, either end-level differences or an order effect using balanced end-level stimuli can explain an overestimation of loudness change for up-ramps. A shortcoming of the current paired-stimulus paradigm is that it is impossible to disentangle the effects of end-level differences and paired-stimulus confounds in global judgments of loudness change. Therefore, Experiment 3 used single rather than paired stimulus trials.

Experiment 3: Effect of End-Level and Intensity Region in a Single-Stimulus Paradigm

In Experiment 3 and using single stimulus presentations, balanced end-level comparisons were made in the analysis between 50−70-dB SPL up-ramps and 90−70-dB SPL down-ramps, without the confounding influence of order effects when balanced end-level paired stimuli are presented. A significant difference in loudness change between up-ramps and down-ramps in balanced end-level comparisons using a single-stimulus paradigm cannot be explained by end-level biases, nor can they be explained by the inherent artifacts of paired stimulus presentations. Furthermore, the end-level effect on up-ramp overestimation between DL and DH conditions reported in Experiment 2 was further investigated in Experiment 3 using single stimulus presentations.

Experiment 3 incorporated a 3 × 2 × 2 × 2 within-subjects factorial design, with three levels of sweep direction (up-ramp, down-ramp, no-change), two levels of timbre (vowel, violin), two levels of duration (1.8 s, 3.6 s), and two levels of sweep region (DL: 50−70-dB SPL, DH: 70−90-dB SPL). Following Neuhoff's (1998) single-stimulus paradigm, the dependent variable was the loudness change of each stimulus, measured by a revised version of the VAS. Following from Experiment 2, it was hypothesized that (a) up-ramps are overestimated in loudness change, relative to down-ramps; (b) loudness change for up-ramps is greater in DH conditions, relative to up-ramps in DL conditions; and (c) up-ramps are overestimated in loudness change, relative to down-ramps, in balanced end-level conditions (e.g., DL up-ramps vs. DH down-ramps).

Method

Participants

The sample consisted of 34 adult participants (25 females and 9 males; *M* = 20.88 years, *SD* = 3.79, range = 18−31 years) recruited from the University of Western Sydney and who did not participate in Experiments 1 or 2. All reported normal hearing. Nine participants had received minimal individual musical training (*M* = 1.17 years, *SD* = .43, range = .5−2 years).

Stimuli and Equipment

Equipment and stimulus generation was identical to Experiment 2, but this time stimuli were not combined into pairs.

Procedure

The procedure closely followed Experiments 1 and 2. However, participants were asked to judge the amount of loudness change within each single stimulus and respond using a revised VAS, where one end of the bipolar scale indicated "No-Change" and the other a "Large-Change," with corresponding scores of 0 and 50, respectively. The order of the bipolar anchors on the VAS was reversed for every other participant to distribute any response bias towards a particular end of the scale. Six practice stimuli were first presented to participants, followed by six blocks of 20 randomized experimental trials (total of 120). The experiment took approximately 20 minutes.

Results

All statistical comparisons were within-subjects contrasts (α = .05) with partial eta squared (η_p^2) as a measure of effect size (Cohen, 1973). It was first hypothesized that up-ramps are overestimated in loudness change, relative to down-ramps in all conditions. This hypothesis was supported with *p* values equal to or less than .001 and effect sizes of $\eta_p^2 = .31$ or above. As can be seen in Figure 7, loudness change for up-ramps was significantly greater than down-ramps in each intensity region, stimulus duration and timbre conditions.

Second, it was hypothesized that loudness change for up-ramps is greater in DH conditions, relative to up-ramps in DL conditions. This hypothesis was supported. The mean loudness change in VAS

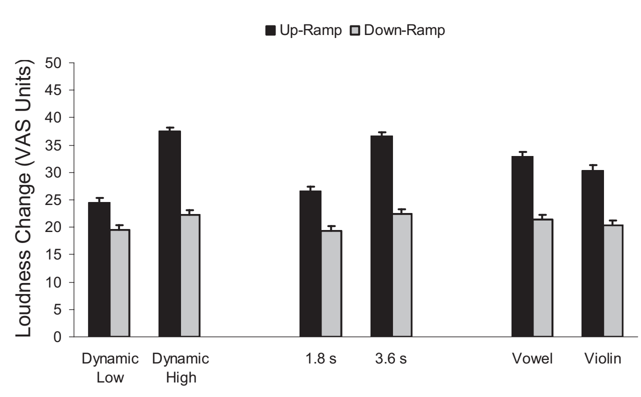

Figure 7. Loudness change in VAS units between up-ramps and down-ramps in Experiment 3. Zero refers to no change in loudness, and 50 refers to a large change in loudness. Error bars represent standard error of the mean. DL = dynamic low (50−70/70−50-dB SPL); DH = dynamic high (70−90/90−70-dB SPL).

units for DH up-ramp stimuli (M = 37.48, SD = 6.70) was significantly greater than the mean loudness change for DL up-ramp stimuli (M = 24.50, SD = 7.48), $F(1, 33) = 203.45$, $p < .001$, $\eta_p^2 = .59$.

Finally, it was hypothesized that up-ramps are overestimated in loudness change, relative to down-ramps, in balanced end-level comparisons. The means associated with planned contrasts that compared DL up-ramps with DH down-ramps are shown in Figure 8. With balanced end-levels of 70-dB SPL, loudness change for vowel 3.6-s DL up-ramps (M = 31.83, SD = 7.05) was significantly greater than vowel 3.6-s DH down-ramps (M = 21.63, SD = 10.22), $F(1, 33) = 29.24$, $p < .001$, $\eta_p^2 = .47$. Furthermore, loudness change for violin 3.6-s DL up-ramps (M = 27.13, SD = 8.07) was significantly greater than violin 3.6-s DH down-ramps (M = 23.39, SD = 9.25), $F(1, 33) = 7.32$, $p < .05$, $\eta_p^2 = .18$.

Loudness change did not significantly differ between vowel 1.8-s DL up-ramps (M = 20.64, SD = 8.72) and vowel 1.8-s DH down-ramps (M = 21.12, SD = 8.79), $F(1, 33) = .12$, $p > .05$, $\eta_p^2 = .00$. In the direction opposite to our prediction, loudness change in response to violin 1.8-s DH down-ramps (M = 21.40, SD = 9.55) was significantly greater than violin 1.8-s DL up-ramps (M = 18.39, SD = 9.32), $F(1, 33) = 6.11$, $p < .05$, $\eta_p^2 = .16$. Therefore, the hypothesis was supported in the 3.6-s conditions only.

Discussion

Experiment 3 investigated intensity dynamics and global judgments of loudness change using a single-stimulus paradigm that removed artifacts associated with paired-stimulus presentations. It was first hypothesized that up-ramps are overestimated in loudness change, relative to down-ramps in all conditions. This broad hypothesis was supported, but it should be noted that by investigating this hypothesis, up-ramps and down-ramps were not balanced in terms of end-levels.

Second, loudness change for up-ramps significantly increased as a function of the intensity region of each dynamic sweep. This result exemplifies the influence of end-level differences on up-ramp perception: each dynamic intensity stimulus in Experiments 2 and 3 covered a 20-dB SPL range, but loudness change in response to up-ramps significantly increased as a function of intensity region and, more specifically, as up-ramp end-level increased.

In addition to effects of end level on up-ramp perception, recency may still bias judgments of loudness change toward the end-level of each item (Susini et al., 2007). Up-ramps may therefore be overestimated in loudness change, relative to down-ramps because up-ramps end 20-dB SPL higher than down-ramps, regardless of whether the intensity sweep covers the DL or DH regions. Planned contrasts in Experiment 3 compared balanced end-level up-ramps and down-ramps in the analysis to control for a response bias based on the most recent intensity level of each stimulus. Therefore, it was hypothesized that up-ramps change more in loudness, relative to down-ramps, when both stimuli end on 70-dB SPL. This hypothesis was partially supported. In balanced end-level conditions, loudness change was greater for up-ramps, relative to down-ramps, when stimulus duration was 3.6-s. This result cannot be explained by an end-level recency effect or paired-stimulus artifacts.

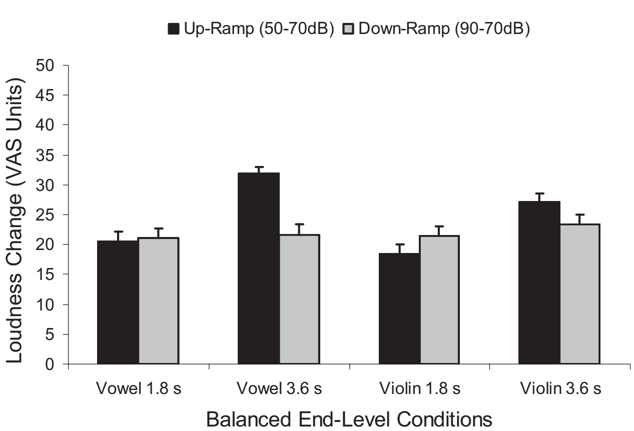

Figure 8. Loudness change in VAS units between 50−70-dB up-ramp and 90−70-dB down-ramp balanced end-level contrasts in Experiment 3. Zero represents no change in loudness, and 50 represents a relatively large change in loudness. Error bars represent standard error of the mean.

General Discussion

Three experiments used single- and paired-stimulus paradigms to investigate global judgments of loudness change in response to dynamic acoustic intensity. In its simplest form, the overestimation of loudness change for up-ramps, relative to down-ramps was recovered with vowel and violin timbres of increasing complexity and duration. In a musical context, this perceptual overestimation may elicit a greater perceived sonic change in response to gradual and extended crescendos, relative to short, abrupt diminuendos (Huron, 1991, 1992)—the classic ramp-archetype from musicological score-based analyses—and may interact in music perception with the correlation between intensity, physiological arousal (Krumhansl, 1997), and heightened emotion (Schubert, 2004). Such an experience may be a fundamental response to an array of musical compositions across cultures. As the dynamic characteristics have been discussed with respect to Western tonal music (Huron, 1990, 1991, 1992), the proposition of empirically investigating this hypothesis in increasingly complex musical contexts and across culturally diverse styles of music awaits further study.

From a psychophysical perspective, Teghtsoonian et al. (2005, p. 705) state that "any attempt to determine the effects of sweep direction on judged change that relies on a single combination of sweep size and sweep location cannot reveal the relative contributions of sweep location and end-level in determining those judgments." The present study has begun to demonstrate how these methodological factors influence perceived loudness change in response to acoustic intensity dynamics. The implementation of important control conditions and the related analyses have shown that the overestimation of loudness change in response to up-ramp stimuli is significantly affected by (a) the duration of each stimulus presentation; (b) the intensity region covered in each dynamic sweep; (c) the difference in end-level between up-ramps and down-ramps; and (d) the order in which paired up-ramps and down-ramps are presented. Therefore, the assumption of an adaptive perceptual bias to rising intensities derived, in part, from global judgments of loudness change, is premature. Nevertheless, differences in perception of up-ramps and down-ramps in the present study remain to be explained.

Sensory and Cognitive Mechanisms: Masking and Memory

Following from work investigating the temporal masking patterns of up-ramp and down-ramp stimuli (e.g., DiGiovanni & Schlauch, 2007; Ries et al., 2008), the overestimation of loudness change that remains in response to up-ramps after methodological artifacts are controlled is partially explained by a simple sensory mechanism. However, sensory mechanisms are most applicable to short stimuli over micro-durations under 1 s (e.g., Stecker & Hafter, 2000). DiGiovanni and Schlauch (2007) used 50-ms and 500-ms up-ramps and down-ramps with unbalanced end-levels (38−80-dB and 80−38-dB SPL) to measure the temporal masking patterns and reported significantly longer post-stimulus neural persistence in response to up-ramp stimuli, relative to down-ramps. It is not surprising that a difference in temporal masking patterns exists between a 42-dB difference in end-level. Investigating temporal masking using up-ramps and down-ramps that cover the intensity sweep regions used in Experiment 2 would elucidate further the relative contribution of a sensory mechanism under balanced end-level conditions. Recently Ries et al. (2008) investigated temporal-masking patterns in response to 10-ms, 50-ms, and 500-ms dynamic stimuli and discuss differences in up-ramp/down-ramp loudness and subjective duration in the context of neural persistence, overshoot, and adaptation. Differences in temporal masking (and auditory threshold shifts from forward masking in particular) in response to dynamic stimuli may be attributed to simple sensory adaptation in the auditory nerve or the persistence of neural excitation at higher levels (Oxenham, 2001).

Modeling peripheral auditory coding of short dynamic stimuli can help ascertain sensory mechanisms involved in perceptual asymmetries; for example, the auditory image model (Irino & Patterson, 1996; Patterson, 1994a, 1994b; Patterson & Irino, 1998). However, as dynamic stimuli extend into macro-durations of approximately 1 s and beyond (Teghtsoonian et al., 2005), cognitive aspects play a greater role. Memory is implicated in dynamic intensity perception over macro durations. Susini et al. (2002, 2007) argue that recency in memory biases global judgments of loudness toward the latter part of a sound sequence. In the context of the paired-stimulus paradigm used in Experiments 1 and 2, recency may not only be dependent on end-level differences. A simpler interpretation of recency is that judgments of loudness change are biased toward the second sound. As loudness adaptation is likely to play a small but significant part in loudness perception at the 50-dB SPL level, this interpretation of recency could help explain part of the "decruitment" effect in the up-ramp/down-ramp DL conditions. It is likely that a general bias of responding to the second sound in a paired-stimulus sequence interacts with a weak loudness adaptation mechanism. This interaction would not affect DH conditions because the intensity sweep region is too high.

Nevertheless, an interpretation of recency based on the second sound in each pair does not apply to Experiment 3, as stimuli were presented in isolation from one another. Furthermore, if we consider recency as a bias only to end-level differences, the significant dynamic balanced up-ramp overestimation in Experiment 3 is still not explained. Recency can apply to each isolated stimulus as a whole. For example, if we consider recency as an integration of a stimulus sequence over a decreasing window of approximately 10 s, then the onset of the up-ramp in global judgments of loudness change will implicate recency when stimuli cover macro-durations, as do the 1.8-s and 3.6-s conditions in the present study. Furthermore, if we consider primacy *and* recency, it is plausible that up-ramps change more in loudness than down-ramps under balanced end-level conditions because the up-ramp becomes louder as it is played through time, relative to its onset intensity level. On the other hand, down-ramps become softer over their duration relative to their onset intensity level, and thus may be perceived to change less. In other words, to consider recency as relevant to only end-level effects is restricting the role of memory. Primacy and recency could help explain a cognitive aspect of the overestimation of loudness change for up-ramps when stimuli cover macro-durations and when methodological artifacts are controlled. Future studies investigating global judgments of loudness change and the role of primacy/recency in memory could present onset and offset intensity levels, with silence as an alternative to the continuous intensity change normally used (e.g., Canévet & Scharf, 1990; Schlauch, 1992). This would determine whether the continuous aspect of intensity change is important, or whether onset and offset levels suffice.

Balanced End-Levels, Perceived Motion, and Loudness Change

Neuhoff (2001, Experiment 2) provides evidence of a bias to approaching stimuli (increasing intensity) under free-field "balanced end-level" conditions. When approaching and receding stimuli (characterized by intensity increase and decrease, respectively) end in the same point (in space), approaching stimuli are perceived to be closer than they really are. Recall that 3.6-s up-ramps were perceived to change more in loudness than 3.6-s down-ramps in Experiment 3 when end-levels were identical (70-dB SPL). These results are evidence from two listening contexts of a perceptual overestimation in response to approaching/up-ramp versus receding/down-ramp stimuli without an end-level artifact. In terms of loudness change, Neuhoff (2001) is concerned with how well judgments of loudness change can be used to infer the distance of approaching and receding sound sources to provide support for evolutionary claims; a fact that may reconcile the controversy surrounding loudness change methods and conclusions drawn (see Canévet, Scharf, Schlauch, Teghtsoonian, & Teghtsoonian, 1999; Neuhoff, 1999 for a brief overview). This is because Teghtsoonian et al. (2005) are primarily concerned with the effect of methods on conclusions regarding loudness perception per se.

The methodological controversy regarding loudness change research needs to give way to new accounts of dynamic intensity perception. It has been shown here that global judgments of loudness change reveal significant findings that are not bound by procedural artifacts. Experiments with converging methods that include magnitude estimation, continuous on-line responses, global judgments of loudness and loudness change, are appropriate in future psychophysical investigations of dynamic stimuli.

Conclusion

The present study has provided new understanding of the intimate relationship between temporal and dynamic aspects of acoustic intensity perception when measured from direct, global judgments of loudness change. These data are valid and reliable when investigated with the appropriate experimental rigor and methodological control exemplified herein. As a result, any explanation or theory that advocates the perceptual salience of dynamic sounds that rise in intensity must acknowledge and control for methodological factors such as stimulus duration, end-level differences and intensity regions within the chosen paradigm, enabling the relative influence of cognitive and sensory mechanisms that underpin the perception of dynamic auditory stimuli to be scrutinized with clarity.

References

Askenfelt, A. (1991). Voices and strings: Close cousins or not? In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Proceedings of Music, Language, Speech and Brain: An International Symposium.* Wenner-Gren Center, Stockholm, 5–8 September, 1990.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17,* 43–64.

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. I. E. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46,* 337–349.

Bregman, A. S. (1999). *Auditory scene analysis: The perceptual organization of sound* (2nd ed.). Cambridge, MA: MIT Press.

Cabrera, D. (1999). The size of sound: Auditory volume reassessed. *MikroPolyphonie,* 5, Retrieved from http://pandora.nla.gov.au/pan/10054/20011007–20010000/farben.latrobe.edu.au/mikropol/volume20011005/cabrera_d/Volume.html

Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America, 88,* 2136–2142.

Canévet, G., Scharf, B., Schlauch, R. S., Teghtsoonian, M., & Teghtsoonian, R. (1999). Perception of changes in loudness. *Nature, 398,* 673–673.

Canévet, G., Teghtsoonian, R., & Teghtsoonian, M. (2003). A comparison of loudness change in signals that continuously rise or fall in amplitude. *Acta Acustica united with Acustica, 89,* 339–345.

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

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33,* 107–112.

DiGiovanni, J. J., & Schlauch, R. S. (2007). Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecological Psychology, 19,* 239–264.

Fant, G. (1970). *Acoustic theory of speech production.* The Hague: Mouton Publishers.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1,* 225–241.

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences, 99,* 15755–15757.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50,* 331–342.

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception and Psychophysics, 68,* 1382–1392.

Grimm, G., Hohmann, V., & Verhey, J. L. (2002). Loudness of fluctuating sounds. *Acta Acustica united with Acustica, 88,* 359–368.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.

Huron, D. (1990). Crescendo/diminuendo asymmetries in Beethoven's piano sonatas. *Music Perception, 7,* 395–402.

Huron, D. (1991). The ramp archetype: A score-based study of musical dynamics in 14 piano composers. *Psychology of Music, 19,* 33–45.

Huron, D. (1992). The ramp archetype and the maintenance of passive auditory attention. *Music Perception, 10,* 83–91.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA MIT Press.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23,* 319–330.

Irino, T., & Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America, 99,* 2316–2331.

Jahnke, J. C. (1963). Serial position effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 2,* 284–287.

Kellaris, J. J., & Kent, R. J. (1993). An exploratory investigation of responses elicited by music varying in tempo, tonality, and texture. *Journal of Consumer Psychology, 2,* 381–401.

Kendall, R. A. (2000). Music Experiment Development System (MEDS) [Computer Software] Retrieved from http://www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67,* 971–995.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51,* 336–353.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the World's languages.* Oxford, UK: Blackwell Publishers.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics, 48,* 169–178.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). Boston: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research, 188,* 70–88.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society, 45,* 224–240.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395,* 123–124.

Neuhoff, J. G. (1999). Perception of changes in loudness': Reply. *Nature, 398,* 673–674.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13,* 87–110.

Oxenham, A. J. (2001). Forward masking: Adaptation or integration? *Journal of the Acoustical Society of America, 109,* 732–741.

Patterson, R. D. (1994a). The sound of a sinusoid: Spectral models. *Journal of the Acoustical Society of America, 96,* 1409–1418.

Patterson, R. D. (1994b). The sound of a sinusoid: Time-interval models. *Journal of the Acoustical Society of America, 96,* 1419–1428.

Patterson, R. D., & Irino, T. (1998). Modeling temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America, 104,* 2967–2979.

Ries, D. T., Schlauch, R. S., & DiGiovanni, J. J. (2008). The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *Journal of the Acoustical Society of America, 124,* 3772–3783.

Rossing, T. D., Moore, F. R., & Wheeler, P. A. (2002). *The science of sound* (3rd ed.). San Francisco: Addison Wesley.

Scharf, B. (1983). Loudness adaptation. In J. V. Tobias & E. D. Schubert (Eds.), *Hearing research and theory* (Vol. 2). Orlando: Academic Press.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331–346.

Schlauch, R. S. (1992). A cognitive influence on the loudness of tones that change continuously in level. *Journal of the Acoustical Society of America, 92,* 758–765.

Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *Journal of the Acoustical Society of America, 109,* 2880–2887.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception, 21,* 561–585.

Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. In Suk-Won-Yi (Ed.), *Music, mind, and science* (pp. 298–352). Seoul: National University Press.

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., et al. (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12,* 2147–2151.

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

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America, 107,* 3358–3368.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. *American Journal of Psychology, 69,* 1–25.

Susini, P., McAdams, S., & Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica united with Acustica, 88,* 536–548.

Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica united with Acustica, 93,* 623–631.

Tajadura-Jiménez, A., Väljamäe, A., & Vastfjall, D. (2008). *Emotional bias for the perception of rising tones.* Paper presented at the Proceedings of Acoustics ’08, Paris, June 29–July 4, 2008.

Teghtsoonian, R., Teghtsoonian, M., & Canévet, G. (2000). The perception of waning signals: Decruitment in loudness and perceived size. *Perception & Psychophysics, 62,* 637–646.

Teghtsoonian, R., Teghtsoonian, M., & Canévet, G. (2005). Sweep-induced acceleration in loudness change and the “bias for rising intensities”. *Perception & Psychophysics, 67,* 699–711.

Zhang, C., & Zeng, F. G. (1997). Loudness of dynamic stimuli in acoustic and electric hearing. *Journal of the Acoustical Society of America, 102,* 2925–2934.

Zwicker, E. (1977). Procedure for calculating loudness of temporally variable sounds. *Journal of the Acoustical Society of America, 62,* 675–682.

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. *Psychological Review, 72,* 3–26.

Received November 20, 2008
Revision received June 4, 2009
Accepted September 8, 2009 ■