32 Auditory Scene Analysis: A Prerequisite for Loudness Perception

Nicolas Grimault[1], Stephen McAdams[2], and Jont B. Allen[3]

1 Introduction

Auditory scene analysis (ASA) refers to the ability of the auditory system to organize perceptually a sound mixture into distinct auditory streams, each stream ideally corresponding to a single sound source (see Bregman 1990 for a review). Level cues can be used to promote perceptual segregation of auditory objects. As such, temporal sequences of pure tone bursts with the same frequencies but with alternating levels can be perceived as segregated as soon as a level difference is introduced between the tones (van Noorden 1975). The inverse relation, i.e., the dependency of loudness on grouping processes, has been much less described in the literature, and the extent to which ASA processes can influence loudness perception remains largely undetermined. Jeng (1992) measured the relative loudness of two distinct sounds (speech and a simulated jack-hammer sound). She found that the total loudness did not follow the Zwicker power-spectrum loudness model (Zwicker and Fastl 1990). This is similar to the situation of a masking noise on speech. As shown by Fletcher (Fletcher and Munson 1933, 1937; Fletcher 1938), speech can be masked by noise, and as the noise level is increased, the loudness of the speech is decreased. From the spectral point of view, the total loudness should be found by summing up all the specific loudness components. In fact, this is not the case, because of grouping. McAdams et al. (1998) provided an argument supporting the assumption that the loudness of an auditory event can be influenced by the perceptual organization. The stimuli used in their experiments were alternating sequences of two identical bursts with no silent gap between them and played at different levels, well above threshold. According to Warren et al. (1972) and van Noorden (1975), these sequences lead to the perception of a continuous sound upon which is superimposed an additional intermittent stream of bursts. This phenomenon suggests that the

[1]UMR 5020, CNRS, Université Claude Bernard Lyon 1, IFR19, Institut fédératif des neurosciences de Lyon, Lyon, France, nicolas.grimault@olfac.univ-lyon1.fr
[2]CIRMMT, Schulich School of Music, McGill University, Montréal, Québéc, Canada H3A 1E3, smc@music.mcgill.ca
[3]Beckman Inst., Urbana, IL 61801, USA, jontalle@uiuc.edu

Hearing – From Sensory Processing to Perception
B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Eds.)

higher level portion is interpreted by the auditory system as being composed of two parts: the continuation of the preceding sound and a new superimposed sound, each with its own loudness. The results of this experiment indicate that the loudness of the intermittent stream can be influenced by up to about 12 dB by this auditory continuity illusion. McAdams et al. (1998) discussed these results in the light of several loudness models and subtractive mechanisms (Warren 1999).

The experiment presented here is a continuation of such investigations between grouping and loudness, across, rather within, equivalent rectangular bandwidths. Our intuitive assumption is that a loudness value must always be associated with an auditory event (Allen 1996). As suggested by the Jeng experiment, two acoustic components belonging to distinct auditory objects do not contribute to a single loudness value. None of the loudness models described in the literature (see Plack and Carlyon 1995 for a review) takes into account the perceptual organization of the stimuli. It follows that the experimental data of McAdams et al. (1998) cannot be predicted by a subtraction mechanism, operating in sone units, that would occur subsequent to this loudness computation. These authors explored the limitations of simple pressure and power models. Predictably, these "linear" models accounted for their data better than the classical loudness models since the stimulus components all lie within equivalent rectangular bandwidths. Altogether, these results are consistent with the hypotheses that first, within a equivalent rectangular bandwidth, pressure must sum, and second, loudness is computed subsequent to auditory grouping mechanisms and consequently depends on these mechanisms. The experimental procedures used only involved the segregation of simultaneous auditory events with similar spectral properties in each equivalent rectangular bandwidth. Given the nature of the stimulus, this experimental design does not address how remote spectral components can lead to a single (or multiple) loudness sensation(s) that depend(s) on their grouping status, given that within a equivalent rectangular bandwidth the pressure sums, whereas across equivalent rectangular bandwidths, loudness sums.

The goal of the present study is to investigate how ASA processes can influence loudness summation across remote equivalent rectangular bandwidths. In other words, this experiment has been specifically designed to check for a possible effect of auditory grouping on the loudness additivity law, as initially defined by Fletcher and Munson (1933); see also Allen (1996). The perceptual grouping status of two simultaneous tonal components has been shown in the literature to be strongly influenced by the spectral relationships they share (spectral regularities, harmonicity, etc.), and by the temporal context in which they are presented (Bregman and Tougas 1989). Thus, the tonal components of a two-tone complex, that would be perceived as integrated, will be heard individually once a tonal context, which forces the segregation, is introduced (Bregman and Tougas 1989). It is known that a slight asynchrony between the onsets of the components will lead to segregation

(Bregman and Pinker 1978). In the present experiment, asynchrony and temporal context was used to promote either integration or segregation across conditions, while the loudness level of the components was measured.

2 Method

2.1 Subjects

Twenty subjects (10 males and 10 females, mean age 26.9 years) with no history of hearing disorder participated; 14 subjects participated in the 50 dB level condition, 15 in the 73.7 dB level condition. Of these nine participated in both conditions and one was dropped from the 73.7 dB condition.

2.2 Stimuli and Procedure

Five experimental conditions (C1 to C5), corresponding to the five sequences of bursts shown in Fig. 1, were generated. The burst duration, in all conditions, was 0.2 s including 3-ms linear rise and fall ramps. Bursts were separated by a 0.1-s silent gap except for C2 in which the bursts were alternately separated by 0.1-s or 0.4-s gaps. All sequences of bursts in all conditions were about 45 s long. During the sequence presentation, the subject's task was always to adjust continuously the level of a particular tone (indicated by an arrow in Fig. 1) to match the loudness of either a pure tone (C1, C2, C3 and C5) or a two-tone complex (C4). The adjustment procedure was terminated at the end of the sequence presentation or when the subject pressed a button on a response box. The last adjusted level was used in this procedure as the adjustment result. Five consecutive adjustment procedures were performed and averaged in each condition. The levels of presentation for all non adjustable 1-kHz pure tones were either 50 or 73.7 dB SPL (the reference levels). All other tone levels were either adjustable or fixed experimentally (see below). The start levels for the adjustable tones were chosen at random between 5 and 8 dB above or below the reference levels for each sequence presentation.

C1 and C2 simply consisted of the presentation of successive bursts of 1-kHz pure tones with two different rhythms. The subject's task was to adjust the level of the tones to attain a sequence of constant loudness across events.

C3 consisted of a repeating sequence including a fixed-level 1-kHz pure tone and an adjustable-level 2-kHz pure tone. The empirically determined matching levels obtained in this condition were then used to generate the two-tone complexes used in C4 and C5, independently for each individual. In fact, testing the additivity law as defined by Fletcher and Munson (1933) and Allen (1996) requires two-tone complexes with equally loud components.

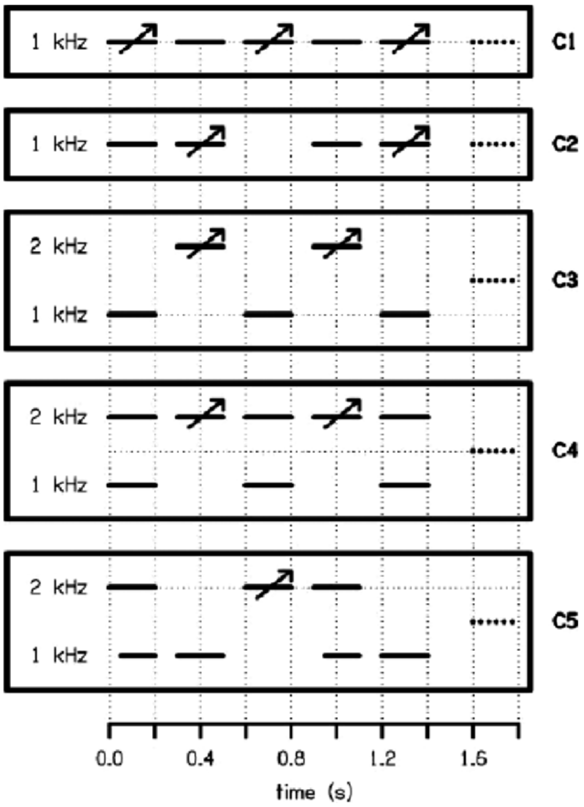

Fig. 1 Schematic time-frequency representations of the first 1.4 seconds of the repeating sequences used in conditions C1 to C5 (see text for details)

C4 consisted of a repeating sequence including a simultaneous two-tone complex with frequencies equal to 1 kHz and 2 kHz at fixed levels determined in C3 and a 2-kHz adjustable-level pure tone. All subjects subjectively reported hearing the complex as integrated before proceeding to the loudness-adjustment task. This condition was designed to match the loudness of a pure tone with those of a perceptually integrated two-tone complex.

C5 consisted of a repeating sequence including: 1) an asynchronous two-tone complex with frequencies equal to 1 kHz (component with 0.05 s asynchrony) and 2 kHz, both having the same loudness, 2) a 1-kHz pure tone, and 3) a 2-kHz adjustable-level pure tone. All subjects but one (who did not perform this condition and was subsequently eliminated from the analyses) subjectively reported hearing the complex as segregated before proceeding to the loudness-adjustment task. The subject's task was to adjust the level of the 2-kHz tones to match the loudness of the 2-kHz component in the complex that was perceptually segregated from the 1-kHz component.

2.3 Apparatus

All stimuli were delivered to the subjects through a Tucker Davis Technology system including an analogue converter (TDT AD1), an anti-aliasing filter (TDT FT6-2), two programmable attenuators (TDT PA4), a mixer (TDT SM3), a headphone buffer (TDT HB6) and a Sennheiser HD 250 headphone. In order to adjust the level of the tones, the subject directly controlled one attenuator (TDT PA4) by using two buttons of a response box (TDT RBOX4/PI2). All signals were calibrated (Larson Davis LD824 system with an AEC101 coupler).

3 Results

The average adjustments in each experimental condition are shown in Fig. 2 for the 73.7 dB SPL (filled squares) and 50 dB SPL (unfilled squares) reference-level conditions. As expected, the adjustments in C1 lead to values very close to the reference levels (74 dB and 50 dB, where 73.7 dB and 50 dB were targeted). As only 9 out of 20 subjects performed both level conditions, two independent statistical analyses (one-way repeated-measures ANOVAs) were performed for each reference level (14 subjects for each reference level) with the experimental condition as the dependent variable. Probabilities are corrected where necessary by the Geisser-Greenhouse epsilon (Geisser and Greenhouse 1958). The experimental condition is significant at the 73.7 dB reference level [$F(4,52) = 5.96$, $p = 0.01$, $\varepsilon GG = 0.44$], but not at 50 dB SPL [$F(4,52) = 2.73$, $p = 0.08$, $\varepsilon GG = 0.54$].

Moreover, a Fisher LSD test applied on the data shows that, at the 73.7 dB reference level, the adjustments in C4 are significantly higher than in all other conditions (C1, C2, C3 and C5). The same statistical test applied on the data at the 50 dB reference level shows that the adjustments in C4 are higher than in C1, C2 and, to a lesser extent, in C5. This pattern of results indicates that all adjustments are about the same (within a level condition) with the notable exception of the higher adjustments in C4. This result is also confirmed by a two-way repeated-measures ANOVA with factors reference level and condition, applied to the adjustments from the nine subjects common to the two level conditions. This analysis reveals a strong effect of level [$F(1,8) = 1024$, $p<0.00001$, $\varepsilon GG = 1$], an effect of condition [$F(4,32) = 6.13$, $p<0.01$, $\varepsilon GG = 0.58$], but no interaction between these factors [$F(4,32)<1$].

4 Discussion

The adjustment procedure used in this experiment seems to be validated by the good correspondence between the mean adjustment values in C1 and the reference levels. Moreover, the consistency between the results in C1 and C2

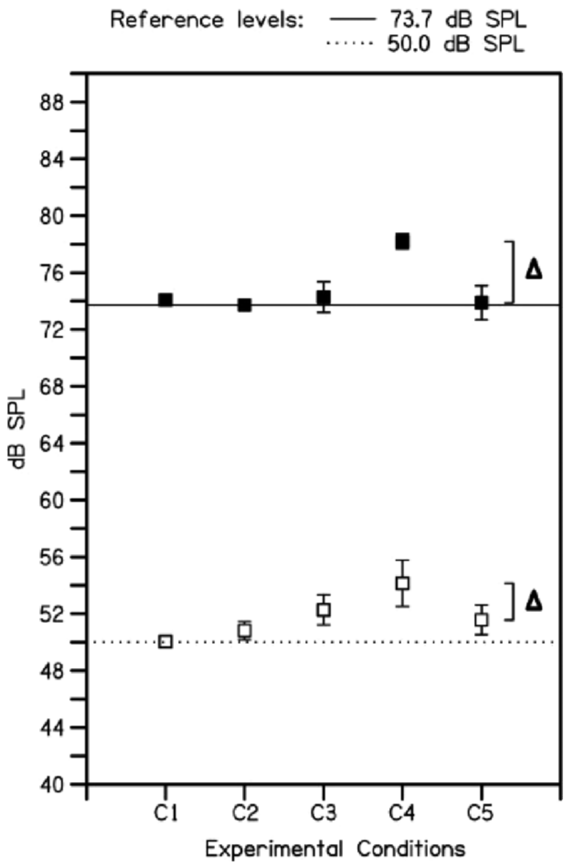

Fig. 2 Mean adjustments across conditions with reference levels equal to either 73.7 dB SPL (*filled squares*) or 50 dB SPL (*unfilled squares*). The reference levels are indicated by the *continuous* (73.7 dB SPL) and *dotted* (50 dB SPL) *lines*. The experimental adjustment differences between C4 and C5 (Δ) are also indicated

indicates a negligible effect of rhythm. The adjustments in C3 indicate that the level difference between 2-kHz and 1-kHz tones that is necessary to obtain equally loud tones, is not significantly different from zero. This result does not compare well with the equal-loudness contours from the literature that generally predict a positive increment. The equal-loudness contours are, however, derived from free field measurements. The headphone presentation, the frequency-response curve of the headphone (Hirahara 1997) and the use of an AEC101 coupler for calibration may have contributed to this difference. This observation makes comparison of the absolute adjustment values obtained in this experiment with those from the literature difficult. The differences across conditions C4 and C5 (Δ in Fig. 2) can, however, be compared with predicted differences from Fletcher and Munson (1933). The adjustment values in C4 are higher on average than those in the other conditions, and in C5 in particular, leading to positive values for Δ. The complex in C5 was reported by all listeners to be perceptually segregated such that the tonal components were then perceived separately. This could presumably explain the higher adjustment values in C4, in which listeners reported the complex tone as being integrated. This result indicates that a single tonal component from a complex can give rise to a particular loudness as soon as it is segregated from the other components (C5), or it can contribute to a global loudness if it is integrated along with the other components into a single auditory object (C4). McAdams et al. (1998) have already demonstrated the influence of ASA processes on within-channel loudness computation processes, where pressure sums. The results from the present experiment therefore provide clear evidence for the influence of ASA processes on across-channel loudness additivity. However, this result must be tempered by the fact that the additivity law (Fletcher and Munson 1933) predicts larger Δ values than the experimental Δ values. The lower-than-predicted Δ values might be explained first by the fact that neither in C4 nor in C5 was the percept clearly integrated or segregated for all subjects. However the individual results were consistent across subjects, and all subjects reported a clear perception of the integrated and segregated events. Thus this explanation would be difficult to support. Second, the lower-than-predicted Δ values might be explained by the short duration of the bursts used in this experiment to promote segregation (0.2 s instead of 1 s in Fletcher and Munson 1933). From the results of Munson (1947), the effect of the reduced burst duration on loudness should not exceed 3 dB, and therefore would not seem to account for the higher observed difference between the additivity law and the data. However, Munson's results might not apply to the loudness of a two-tone complex. Third, the lower-than-predicted Δ values might finally be explained by biases. In fact, in the loudness adjustment procedure, the complex tone was always fixed and the pure tone variable. The other way round has not been tested as recommended by Gabriel et al. (1997). Moreover, the range of starting levels for the adjustable stimulus was not determined experimentally to be symmetric around the final adjusted value. These biases should however apply similarly to C4 and C5 and would not explain the lower-than-predicted Δ values. This therefore requires further experimental analysis involving additional experimental conditions with multi-component complex tones. This study, along with McAdams et al. (1998), suggests that the auditory stimulus representation, at the stage where ASA processes occur, must occur prior to loudness computation. Indeed, these studies demonstrate that loudness perception is strongly influenced by both within-channel and across-channel ASA mechanisms, because grouping processes have been shown to affect loudness perception in the case of homophonic continuity (McAdams et al. 1998) and in more complex sounds in the current study. We conclude that loudness models must be revised to account for this large ASA effect that has been largely underestimated.

Acknowledgments. We thank Brian Moore for helpful comments on a previous version of this manuscript.

References

Allen JB (1996) Harvey Fletcher's role in the creation of communication acoustics. J Acoust Soc Am 99:1825–1839

Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. MIT, Cambridge, MA

Bregman AS, Pinker S (1978) Auditory streaming and the building of timbre. Can J Psychol 32:19–31

Bregman AS, Tougas Y (1989) Propagation of constraints in auditory organization. Percept Psychophys 46:395–396

Fletcher H (1938) Loudness, masking and their relation to the hearing process and the problem of noise measurement. J Acoust Soc Am 9:275–293

Fletcher H, Munson WA (1933) Loudness, its definition, measurement and calculation. J Acoust Soc Am 5:82–108

Fletcher H, Munson WA (1937) Relation between loudness and masking. J Acoust Soc Am 9:1–10

Gabriel B, Kollmeier B, Mellert V (1997) Influence of individual listener, measurement room and choice of test-tone levels on the shape of equal-loudness level contours. Acustica - Acta Acustica 83:670–683

Geisser S, Greenhouse SW (1958) An extension of Box's results on the use of the F distribution in multivariate analysis. Ann Math Stat 29:885–891

Hirahara T (1997) Physical characteristics of headphones used in psychoacoustical experiments. J Acoust Soc Jpn 798–806

Jeng PS (1992) Loudness prediction using a physiologically based auditory model. Unpublished PhD thesis, CUNY, New-York

McAdams S, Botte M-C, Drake C (1998) Auditory continuity and loudness computation. J Acoust Soc Am 103:1580–1591

Munson WA (1947) The growth of auditory sensation. J Acoust Soc Am 19:584–591

Plack CJ, Carlyon RP (1995) Loudness perception and intensity coding. In: Moore BCJ (ed) Hearing. Academic Press, pp 123–159

van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Unpublished Doctoral Dissertation, Technische Hogeschool Eindhovern, Eindhoven, The Netherlands

Warren RM (1999) Auditory perception: a new analysis and synthesis. Cambridge University Press

Warren RM, Obusek CJ, Ackroff JM (1972) Auditory induction: perceptual synthesis of absent sounds, Science 176:1149–1151

Zwicker E, Fastl H (1990) Psychoacoustics – facts and models. Springer, Berlin Heidelberg New York