References

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*, 277-286.

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronic sequence. *Journal of the Acoustical Society of America*, *98*, 2524-2531.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127*, 22-44.

Grondin, S. & Killeen, P. R. (2006). Counting and singing improve the timing of long intervals. In D. Kornbrot, R. Msetifi, &S. Macrae (Eds.). *Fechner Day 2006: Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics* (pp. 193-198). St.Albans, UK: The ISP.

Grondin, S. & Killeen, P. R. (in preparation). Why Galileo sang: Measuring the passage of time by singing and counting.

Grondin, S., & Laforest, M. (2004). Discriminating slow tempo variations in a musical context. *Acoustical Science & Technology, 25*, 159-162.

Grondin, S. & McAuley, J. D. (2007). Crossmodal comparisons of time intervals presented or not presented in sequences. *Poster presented at the 48th Annual Meeting of the Psychonomic Society* (Long Beach, Ca*, nov*).

Grondin, S., Meilleur-Wells, G., & Lachance, R. (1999). When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process by humans. *J of Experimental Psychology: Human Perception and Performance, 25*, 993-1004.

Hinton, S. C., & Rao, S. M. (2004). ''One thousand-*one* … one-thousand-*two* …'': Chronometric counting violates the scalar property in interval timing. *Psychonomic Bulletin & Review*, *11*, 24-30.

Jones, M. R., & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Perception & Psychophysics*, 67, 398-417.Keele, S. W., Nicoletti, R., Ivry, R., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgements? *Psychological Research*, *50*, 251-256.

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review, 94*, 455-468.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time varying events. *Psychological Review*, *106*, 119-159.

McAuley, D. J., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1102-1125.

Miller, N., & McAuley, J. D. (2005). Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Perception & Psychophysics*, *67*, 1150-1160.

Acknowledgement.

This research is supported by a research grant from the Natural Sciences and Engineering Research Council of Canada (simon.grondin@psy.ulaval.ca). I would like to thank Nicolas Bisson for his help in the preparation of this presentation.

RHYTHMS EMERGE FROM THE PERCEPTUAL GROUPING OF ACOUSTIC COMPONENTS

Albert S. Bregman
Psychology Department, McGill University
al.bregmanationmcgill.ca

Abstract

To have a rhythm, one has to have a succession of distinct units. Usually we take the existence of auditory units for granted, because we have created them. We think of them as separated by silences, but there is hardly ever total silence in a complex auditory scene. However, the auditory system has mechanisms that find units within it. A unit not only contains acoustic energy that stretches over time but over the spectrum. Various acoustic factors favour the integration of energy into the same auditory unit. Once formed, these units have to be bound together sequentially to form an auditory stream. To a first approximation, a rhythm emerges within a stream. Although the ideal function of a stream is to represent a particular source of sound, in practice there is not always a one-to-one correspondence between streams and sources. Research and demonstrations will be presented to illustrate these ideas.

The relation between auditory scene analysis (ASA) and rhythm is not unique, but is an instance of a more general rule: "Every perceived quality of sound is affected by perceptual organization". Examples of perceived qualities are position in space, loudness, and pitch.

Perceived spatial position, loudness, and pitch

The spatial position of sounds can be affected by ASA. I will illustrate this by mentioning the unpublished research of two of my students, Yuko Tagami and Melissa Rapoport at McGill.

In both studies, the two ears of the listener received the identical sound (a pure tone binaurally matched in frequency, phase, and intensity). Normally this should be heard as a single centred sound. However, in both these experiments the sound at one ear was longer, starting before and ending after the sound in the other ear. During the period in which they coincided, they were identical in all respects. Nonetheless, due to the asynchrony of onset, they were often heard as two sounds, one at each side of the head (violating the common belief that a single frequency cannot be heard in two places at the same time). In Tagami's research, listeners described the number of sounds they heard and their positions by choosing among a set of visual icons representing alternative percepts. In Rapoport's study, they adjusted the interaural intensity of a binaural reference tone to match the position at which they had heard the shorter of the two tones. Both experiments showed that the faster the rise time of the shorter tone, the more segregated it was from the longer tone and the more it was lateralized to one side. The longer the temporal overlap of the two tones, the more they tended to fuse and be heard as a single centred tone.

Perceived loudness can also be affected by perceptual organization. Suppose a soft steady noise burst, A1, is on for a brief period, and then suddenly is replaced by a louder version of itself, A2, and then the soft sound, A1, returns (all without silences or breaks). In certain conditions, listeners will hear two sounds. One of them is the soft sound A1, which is heard as continuing unchanged during the louder period and continuing after the loud period ends. The loud period (A2) is heard as the occurrence of a second sound, B, superimposed on the unchanging A1. In other words the loud sound, A2, is perceptually interpreted as the sum of an unchanging A1 and a new sound, B. These two are heard as soft sounds whose addition produces the loud sound A2 (Warren, Obusek & Ackroff, 1972; also Bregman, 2000, pp.371-3). So the ASA process causes the listener to interpret the sensory stimulation that would normally come from a single loud sound to be coming from two softer sounds.

Perceived pitch can also be affected by ASA. It is known that mistuning one of the lower harmonics, say the third one, of a complex tone by raising its frequency a little, can raise the perceived pitch of the overall tone. However, perceptual grouping can affect this phenomenon. If the mistuned partial is made to start a bit before the rest of the tone, or end a bit later, it is perceived as a separate sound, and its effects on the pitch of the complex tone will be removed, and the tone will be heard as if the offending partial were absent. (Darwin & Ciocca, 1993). Here we see how the allocation or non-allocation of a partial to a complex tone by the ASA system can affect the perceived pitch

Rhythm: Units and Streams

The effects of ASA processes on auditory rhythms are just another example of the functioning of this system. Rhythms involve a regular repetition of a short sequence of discrete sounds, sometimes thought of as units. Therefore, before one can have rhythmic sound one needs the following: (a) units of sound, (b) Integrated sequences or streams of these units, and (c) awareness of a regular repetition of sub-sequences within these streams. The first two of these lie within the scope of ASA; so I will restrict my discussion to them.

Formation of units.

In order to have rhythms, one has to have a succession of distinct units. In laboratory research and in music, we take the existence of auditory units for granted, because we have created them – for example a series of tones or noise bursts, or musical notes. We think of them as separated by silences, just as the words of a written sentence are separated by blank spaces. But there are hardly ever complete silences in natural environments, especially the urban environments in which most of us now live. However, the auditory system has mechanisms that can take a stretch of sound and form units within it.

Sudden rises in intensity in one or more spectral regions cause the auditory system to start a unit. The unit is ended either when the intensity drops to the starting value, or another unit replaces it. For a detailed analysis of the rules for the formation of sequences of distinct units when the evidence for units overlaps in time, we can refer to the studies of Nakajima and his colleagues and Nakajima's event construction model (e.g., Nakajima et al., 2000) and his demonstrations (Nakajima, 2006).

Units appear to be a focus for perceptual analysis. There is a tendency to form the description of a unit of sound based on the properties of its onset, and to maintain those properties for the duration of the unit. Consider a tone that rises abruptly in intensity and decays slowly. The percussive onset seems to engage the auditory system's analysis of the pitch of the tone. If such tones are temporally reversed, so that the sudden change in intensity comes at the end, they don't seem to have the same vividness of properties. For example listeners were asked to discriminate the order of a rapid overlapping sequence of four such (pure) tones. They were able to do so more accurately when the onsets, rather than the offsets were abrupt. The more abrupt the onset, the easier the task (Bregman, Ahad & Kim, 1994). We seem to physiologically respond to onsets more than to offsets, and this has a important value for the animal that is built this way (it is better to flee when the sound made by a begins than after it ends).

Another study illustrates the importance of units in accurate perception (Crum & Bregman, 2006). Listeners were presented with a 5-s tone that gradually changed in timbre,

and were asked to press a button as soon as they heard a change in timbre. There were three conditions: (*a*) the tone was continuous and unbroken; (*b*) short silences of 20 ms replaced parts of the tone at regular periods, breaking the 5-s tone into a sequence of short units; (*c*) loud noise bursts were inserted instead of silences. In the latter case, the bursts induced the illusion that the tone continued behind them, so that the experience of short units was reduced or eliminated. In Condition *a*, in which discontinuous units were perceived, the change in timbre was detected about 270 ms earlier than in the two perceptually continuous conditions, which did not differ from one another. It seems that due to its focus on onsets. Our auditory system can detect changes in the properties of a succession of units more quickly than in a continuously changing sound.

Sequential integration of units into streams.

An auditory rhythm is more than a sequence of separate units. The units must be perceptually grouped into a coherent sequence, or auditory stream (not all the sound that is present at a given moment will be integrated into the same stream). A rhythm is perceived wholly, or at least most strongly, *within* a stream. Although the ideal function of an auditory stream is to represent a particular source of sound, in some cases the one-to-one correspondence between streams and sources is violated.

An example of this is that rhythms can emerge from the grouping of components from different physical sources. For example, the African amadinda, a twelve-tone xylophone, can be played by two people, each striking the keys in strict alternation with the other. Each player's strikes are of metronomic regularity; yet the listener hears a highly complex rhythm. This occurs because each player is playing a sequence that contains both higher and lower notes. Two auditory streams are formed, a lower one and a higher one, each containing notes from both players. Since each player's pitch sequence is irregular in a planned fashion, the emergent high stream and low stream can contain complex pre-planned rhythms. Thus the sounds from two physical sources (the players) combine to form each stream because of their acoustic similarities. (Bregman & Ahad, 1996, Demonstrations 7-9; Wegner, 1993).

Another example of the within-stream perception of rhythms occurs in polyrhythms. A polyrhythm is formed when two different rhythms are played at the same time, often by the two hands of a single player. An example is where the right hand plays four beats to every three beats of the left hand. Each individual hand plays a regular metronome-like sequence. However, if the sounds coming from each hand have exactly the same qualities (pitch, timbre, etc.), it is virtually impossible for listeners to hear the two regular component rhythms. Instead they hear the irregular cross-rhythm that consists of all the beats forming a single sequence. But when the component rhythms are carried by two notes of different frequencies, the listener can hear the sequence as two simple rhythms being played at the same time. That stream segregation is involved is shown by the fact that the perceptual isolation of the two component rhythms can be strengthened by two factors known to affect stream segregation: increasing the frequency separation of the two tones, and increasing the speed of the sequence (see Experiments reviewed by Bregman 2000 and by Handel, 1984).

Simultaneous integration across frequency

Because rhythm is sensitive to stream segregation, it can be used as a way of studying the factors that promote the fusion of acoustic components. The method is called "rhythmic masking release" (e.g., Turgeon, Bregman & Roberts, 2005). Using repetitions of a short pure tone of a fixed pitch, the experimenters created two stimuli, each with a different rhythm, (A and B) which the listeners had to discriminate. This was quite easy. It is made harder by the placement of distracter tones of the same frequency and duration as the original rhythmdefining (R) tones, at random intervals between the R tones. The Rhythm A and B sequences now sounded like random temporal patterns of tone *f*, and any detection of the embedded rhythm was impossible. To make it easy again, we caused the distracter tones to have a different timbre from the R tones. We did this by adding "flanker" tones, more or less synchronous with the distracter tones, and above and below the latter in frequency. If the flankers fused perceptually with the distracters, the fused tones now had a timbre distinct from the pure-tone quality of the R tones and formed a separate complex-tone stream. Now the original R tones could be heard again as a stream of pure tones. This recovery of discrimination depends on the timbre caused by the fusion of the distractor tones with their and flanker tones; so the ability to discriminate can be used as a measure of the fusion of these tones. Using this method, we found that the most important factors relating the flankers and distractors, at least in the case of the short tones that we used, was how synchronous their onsets were and the duration of their temporal overlap. Harmonic relations between the partials had some influence, but only an extreme degree of spatial separation (dichotic presentation) affected fusion (Turgeon, Bregman & Roberts, 2005).

These examples illustrate the close relation between ASA and rhythm. Rhythms are formed out of units, linked sequentially into auditory stream streams. The processes that form both units and streams are general ones that operate with all sounds. For this reason, the general principles of ASA can help us predict the formation of rhythms and rhythms can be used to study ASA.

References

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA.: MIT Press. (Paperback, 2004)

Bregman, A.S., & Ahad, P. (1996) *Demonstrations of auditory scene analysis: The perceptual organization of sound.* Audio compact disk. (Distributed by MIT Press).

Bregman, A.S., Ahad, P., & Kim, J. (1994). Resetting the pitch analysis system. 2: Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *Journal of the Acoustical Society of America*, *96*, 2694-2703.

Crum, P.A., & Bregman, A.S. (2006). The effects of unit formation on the perception of a changing sound. *Quarterly Journal of Experimental Psychology. 59 (3),* 543-556.

Darwin, C.J., & Ciocca, V. (1993). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, *93 (5)*, 2870-2878.

Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception, 1,* 465-484.

Nakajima, Y. (2006) Demonstrations of the gap transfer illusion, *Acoustical Science & Technology*, 27, 322-324. *http://www.kyushu-id.ac.jp/~ynhome/ENG/Demo/illusions.html*

Nakajima, Y., Sasaki, T., Kanafuka, K., Miyamoto, A., Remijn, G, & ten Hoopen, G. Illusory recouplings of onsets and terminations of glide tone components. *Perception & Psychophysics, 62 (7)*, 1413-1425.

Turgeon, M., Bregman, A.S., & Roberts, B. (2005). Rhythmic masking release: Effects of asynchrony, temporal overlap, harmonic relations, and source separation on cross-spectral grouping. J*ournal of Experimental Psychology: Human Perception & Performance. 31 (5),* 939–953.

Warren, R.M., Obusek, C.J., & Ackroff, J.M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, *176*, 1149-1151.

Wegner, U. (1993). Cognitive aspects of amadinda xylophone music from Buganda: Inherent patterns reconsidered. *Ethnomusicology*, *37*, 201-241.

OBTAINING THE LOUDNESS EXPONENT FROM BINAURAL AUDITORY ADAPTATION DATA

Kenneth H. Norwich and Lisa M. D'Alessandro
Institute of Biomaterials & Biomedical Engineering,
Department of Physiology
University of Toronto
k.norwich@utoronto.ca, l.dalessandro@utoronto.ca

Abstract

The loudness of a stimulus, L, is related to the intensity of that stimulus, φ, *by Stevens' Power Law,* $L = K\varphi^n$, *where K is an arbitrary scaling constant and n is the loudness exponent. The exponent, n, has been measured extensively by psychophysicists usually by magnitude estimation. We devised a new method for obtaining n from experiments using the SDLB (simultaneous dichotic loudness balance) technique. That is, we presented a 6-min continuous tone to a participant's adapting ear. At 1-min intervals, participants adjusted the intensity of the tone in the contralateral control ear until both tones sounded equally loud. We used the control ear, which was otherwise retained in silence, to measure the progressive adaptation in the adapting ear. The value of n may be estimated by dividing the number of decibels of adaptation at 6 minutes (near equilibrium) by the intensity of the continuous tone in dB SL.*

The loudness exponent, *n*, defines the relationship between the intensity of a pure tone (measured, for example, in watt.cm^{-2}) and the loudness of the tone (measured sometimes in units of *sones*, but in all cases dimension-free). In this paper we shall express intensity as a dimensionless ratio, $\varphi / \varphi_{\text{thresh}}$, where φ_{thresh} is the threshold of the listener. $10 \log_{10} (\varphi/\varphi_{\text{thresh}})$ expresses the intensity of a tone as *sensation level* or SL. The equation that explicitly connects loudness with intensity is the power law of sensation, developed by S. S. Stevens:

$$L = k(\varphi/\varphi_{\text{thresh}})^{n}$$

In this equation, *k* is a scaling constant, greater than zero but otherwise arbitrary, that determines the magnitude of *L*. The exponent, *n*, is of paramount importance in psychoacoustics. This exponent compresses the physical variable, intensity ($\varphi/\varphi_{\text{thresh}}$), which extends over a range of about 9 decades in everyday life into a range of only about 3 decades for the psychophysical variable, loudness, *L*. The value of *L* is commonly measured by the method of magnitude estimation. Using this technique, the value of *n* has been found to be about 0.3 for all auditory frequencies between about 400 to 10,000 Hz

It had generally been held that there was no gender difference in the values of *n.* However, research conducted by our group (Sagi, D'Alessandro and Norwich, 2007) suggests that the loudness exponent for females exceeds that for males. In Sagi *et al*. (2007), we introduced techniques for measuring the *relative* values of *n* (females vis-à-vis males) without using classical subjective assessment such as magnitude estimation. Participants, who had been appropriately trained, were required to identify the dB intensity of unknown tones. From the *errors* made by these participants we were able to estimate relative values of *n* over a range of frequencies (125 to 8000 Hz). When measured by this technique, the mean value for *n* over all auditory frequencies we tested was 0.3053 for females and 0.2218 for males. These observed differences in *n* between genders were statistically significant. In the present study, we examine another method of quantifying the loudness exponent, *n*, using an independent set of experiments: studies on auditory adaptation.