WHEN TIME PERCEPTION GOES WRONG

Elvira Pérez and Yoshitaka Nakajima
Department of Acoustic Design, Kyushu University, Shiobaru, Minami-ku, Fukuoka 815, Japan
Perez.elvira@gmail.com   Nakajima@design.kyushu-u.ac.jp

Abstract

Auditory illusions illustrate how the auditory system, by imposing a set of rules and temporal constraints, avoids 'ungrammatical' percepts. To test this hypothesis, we measured the listeners' ability to resolve ungrammatical patterns by creating complex tones with asynchronous offset time. An 'unusual' offset asynchrony was created by prematurely ending only one of the two components. In some conditions, two consecutive tones were perceived. The results showed a task order effect and a tendency to overestimate the perceptual duration of the second tone, which should have been the rest of the complex after one of the components was ended. These auditory illusions illustrate that time perception, with regard to audition, does not correspond to the immediately present acoustic reality, but rather to the outcome of temporal integration.

Auditory Grammar (AG) is a theoretical model for sound organization (Nakajima & Sasaki, 1995). AG is closely related to de Saussure's openings and closing movements for the emission of speech (de Saussure, 1966), and Gestalt perceptual grouping principles (Koffka, 1935) in the formation of auditory streams or events (Handel, 1989; Bregman 1990). Accordingly, the auditory system imposes a *grammar* in which acoustic units or subevents (i.e., onsets, <, fillings, =, offsets, >, and silence, /) are combined and reorganised to create auditory events (e.g., /<=>/). We hypothesise that the termination of a sound has to be perceptually linked to a specific beginning to be perceived as an auditory event. If one component of a complex sound ends prematurely (e.g., ungrammatical construct: <=>=>), the offset of that component will be link to the nearest onset, and the remainder of that sound will have to be reinterpreted as a distinctively new event. The perceptual duration of this second new event will be overestimated because the auditory system has to resolve an ungrammatical pattern by either inserting an onset (e.g., grammatical construct: <=> + <=>) or replacing the asynchronous offset with an onset (e.g., grammatical construct: <=<=>).

We present a novel auditory illusion called *subevent restoration* which strengthens the validity of this grammar.

Method

Participants were asked to perform three different tasks. First, they were asked to just listen all the stimuli once. Then, they were asked to describe the stimuli presented in terms of duration and pitch. To do that, two A4 (210 mm x 297 mm) sheets of sectioned paper (180 m/m x 250 m/m) were given together with a ruler, a pencil, and a rubber eraser. Participants were allowed to listen to the stimulus patterns as many times as necessary and no time limit was given. Finally, participants were asked to perform a magnitude estimation task in which they had to judge the total duration of a standard pure tone, or a part of the asynchronous complex. That is, in *condition 1* participants had to judge the duration of the complex before the asynchronous offset (‘the first part’ in the instructions), and in a *condition 2* they had to judge the duration of the complex after the offset asynchrony (‘the second part’ in the instructions). These conditions were counterbalanced, and the same set of stimulus patterns appeared in both conditions. Stimuli automatically appeared in random order every 10 seconds (on inter-onset interval). Within that 10 seconds participants had to judge the duration of either the first or the second part, and insert a numerical value in a box displayed on a computer interface. Feedback was not provided. There were 40 trials in each block. Each participant performed six blocks (three blocks per condition).

Instructions for the magnitude estimation task. The following instructions were given in both English and Japanese, together with some demonstrations and practice trials: “This is a Magnitude Estimation experiment. You have to estimate the duration of the stimuli. Please, judge the duration of the whole signal when it is a single steady-state tone. Otherwise, judge the duration of the FIRST/SECOND part if it is a non-steady-state tone (pitch changes). Please, insert a numerical value in the box. You can use any numerical value (decimals are OK, but not negative values). The first stimulus pattern will appear automatically, so be ready. The stimuli will appear automatically every 10 seconds. Try not to press replay. After 40 trials you can leave the booth and rest. Thank you.”

Participants. Two males and two females, all with normal hearing. They were 21-24 years old.

Apparatus and stimuli. All the sound patterns (n=37) were generated digitally (16 bit; a sampling frequency of 44100 Hz) and controlled by a computer (Frontier KZFM71/N) with an audio card (E-MU, 0404). The stimulus patterns were presented via a digital-to-analog (D/A) converter (Fostex VC-8); an active low-pass filter at 15 kHz (NF DV8FL); a digital graphic equalizer (Roland RDQ-2031); an amplifier (Stax SRM-313), and headphones (Stax SR-303) to the participant in a soundproof room binaurally. Stimulus patterns were presented at 70 dBA (Fast-peak). The levels were measured with a sound level meter (Node 2072), and an artificial ear (Brüel & Kjær 4153). The stimuli presented comprised a 2-component tone of 500 Hz and 2500 Hz (harmonic), and 499.91 Hz and 2095.55 Hz (inharmonic) respectively. Asynchrony was created by terminating one component before the other in the complex. The asynchronies either in the first or the second component had a duration of 290, 270, 240, 160, 80, 50, or 30 ms. As control we used 1500 Hz steady-state tones of 320, 290, 270, 240, 160, 80, 50, 30, and 20 ms of duration. The onset and the offset of each component were 15 ms shaped separately using a cosine function.

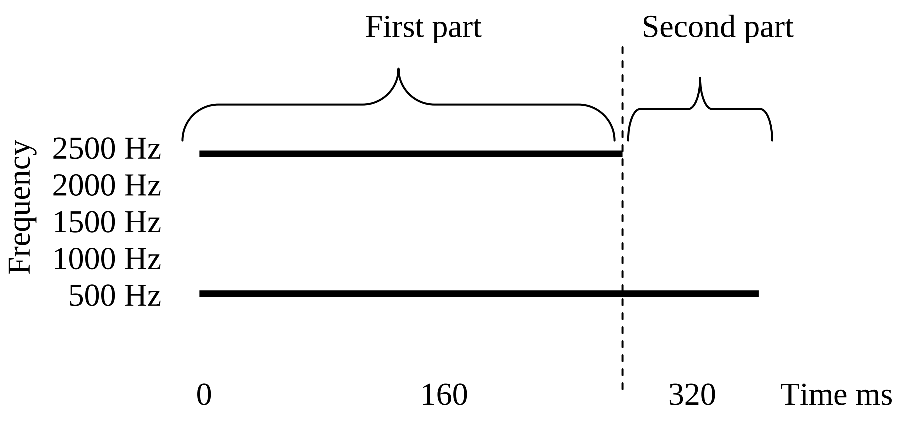

Fig. 1. Schematic representation of a harmonic complex with an offset asynchrony of 80 ms. In condition 1, participants were asked to judge the duration of the first part (160 ms), while in condition 2, participants were asked to judge the duration of the second part (80 ms).

Results and Discussion

Phenomenological data

In general, when participants were asked to describe the percept, there was a very accurate correspondence to the stimulus patterns. Listener F detected offset asynchronies larger or equal to 80 ms. Listener S and E detected offset asynchronies larger or equal to 160 ms, while listener H detected offset asynchronies larger or equal to 240 ms. No main differences were observed when describing the percepts of harmonic and non-harmonic complex.

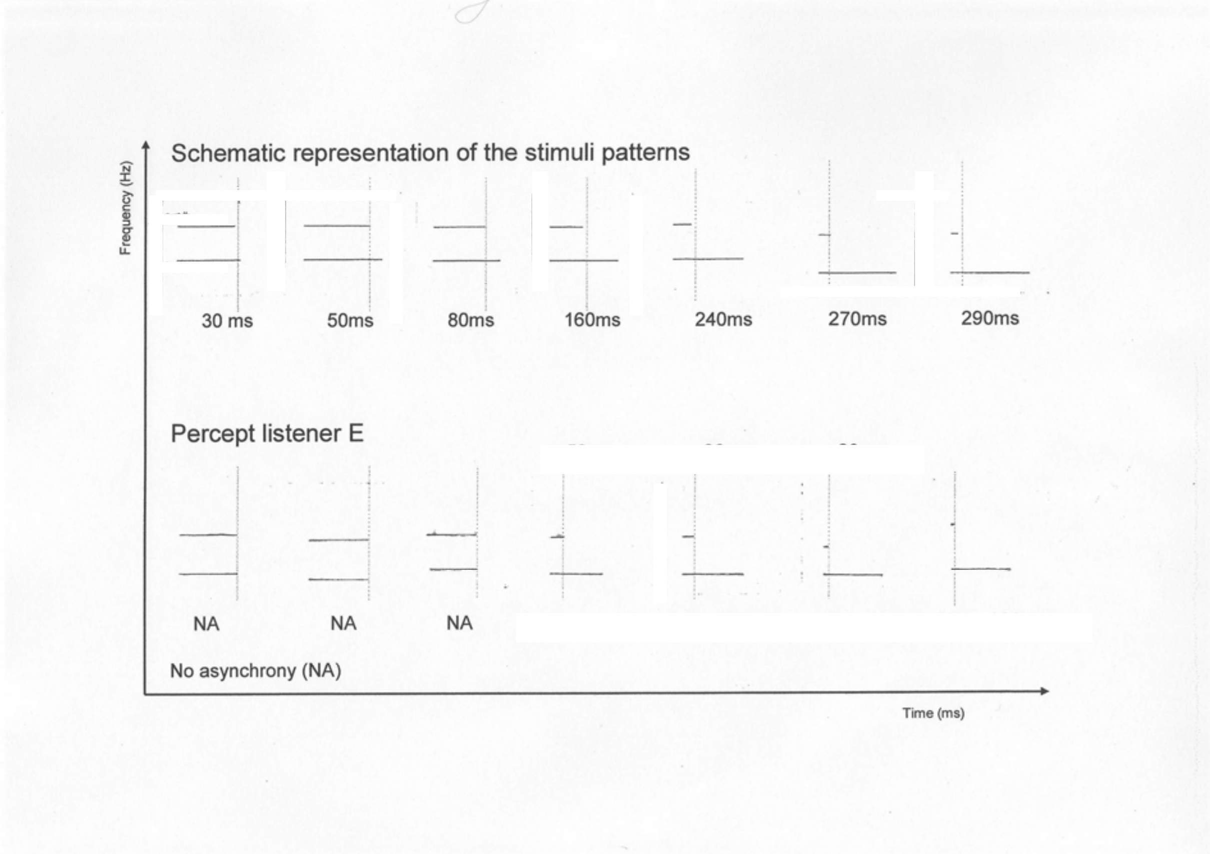

Fig. 2. Comparison of the physical properties of the stimuli with the reported percepts of listener E. The first row is a schematic representation of the harmonic complex. The *y*-axis represents frequency (Hz) and the *x*-axis time (ms). In this example, the second component is the asynchronous one ending 30, 50, 80, 160, 240, 270 or 290 ms before the first component. The second row represents the percepts of one of the participants. It can be observed that the asynchronies of the first three stimuli are not represented (both components end simultaneously). Only when perceiving the fourth stimuli (physical asynchrony = 160 ms), the asynchrony is detected. Larger asynchronies are perceived showing a categorical change.

Magnitude estimation data

To further contrast and corroborate the above results, a magnitude estimation task was performed. Scores were standardized (z scores) by subtracting the estimated mean of the population (μ) from the raw score (x) and dividing the output by the estimated standard deviation of the population (σ).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

The z scores derived from the estimations of the duration of the control tones were compared with those derived from the estimations of the duration of the first and the second part of the complex tones (one-way ANOVA). The results in figure 3, 4 and 5 show the perceived durations for the *control*, *first* part of the complex, and *second* part of the complex. Filled stars denote overestimation, while white stars denote underestimation. Participants S & F first performed the task in which the first part of the complex had to be measured, while participants E & H first performed the task in which the second part of the complex had to be measured. Both sets of data did not statistically differ within groups, but differed between groups, therefore S & F data and E & H data were collapsed respectively. The difference between both groups illustrates an interesting order effect, that is, listeners S & F's strategy to judge the duration of the second part was determined by the first part. These listeners reported after the experiment that involuntarily they were deducing the second part by subtracting the first part from the whole part.

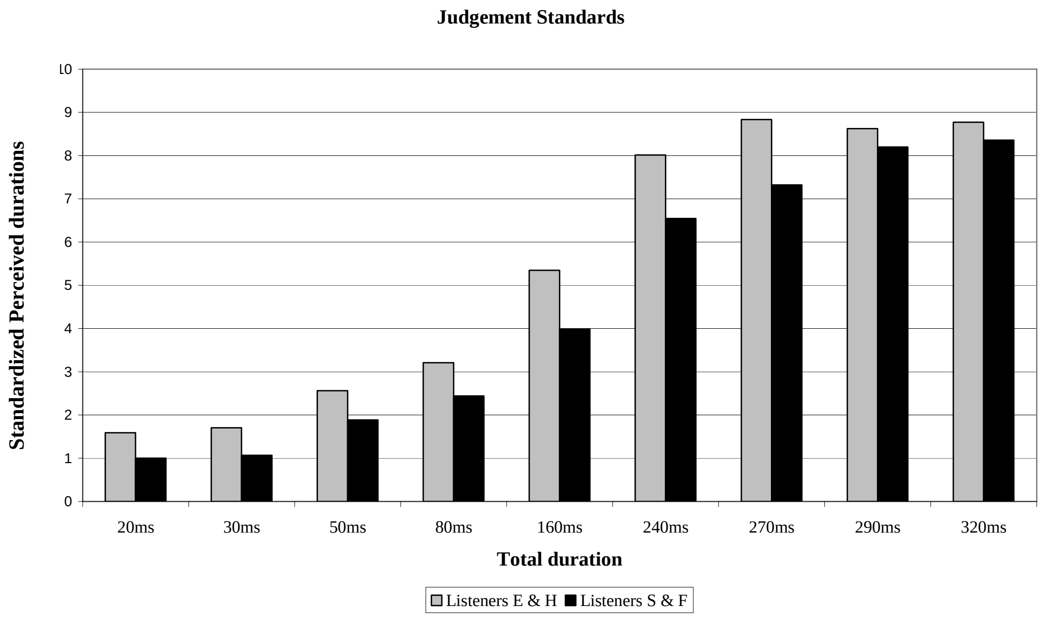

Fig. 3. Results for the control condition in which listeners had to judge the total duration of a 1500 Hz tone. The longer the physical duration is, the longer it is perceived.

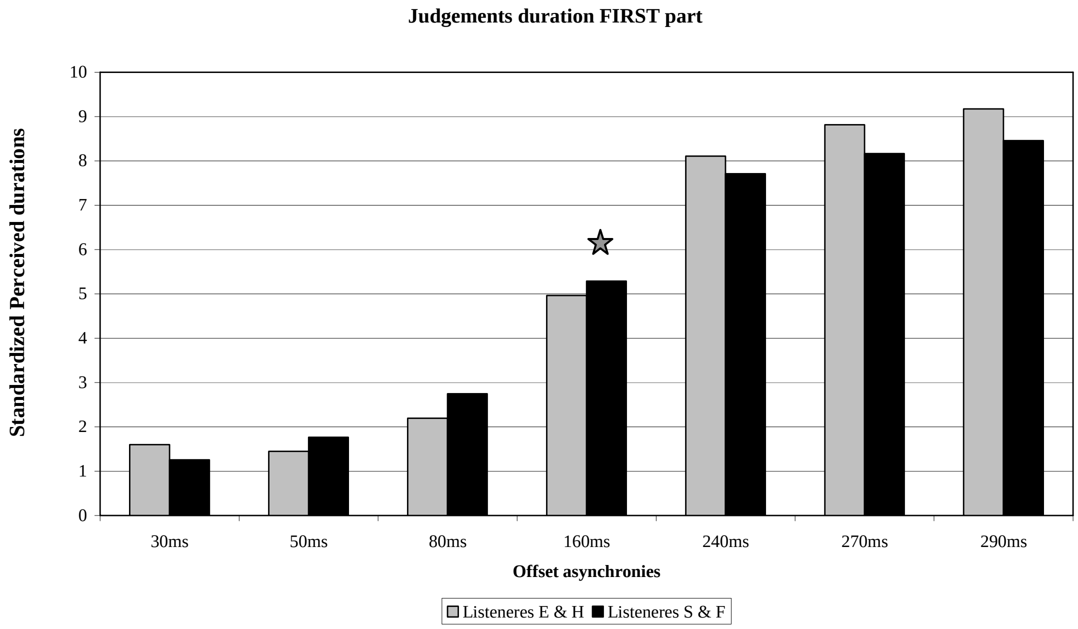

Fig. 4. Results for the judgements of the first part of the complex. When comparing the perceived durations of the standard with the first part, participants S & F seem to overestimate the 160 ms asychrony [F (1,7) = 8.34, $p<0.05$].

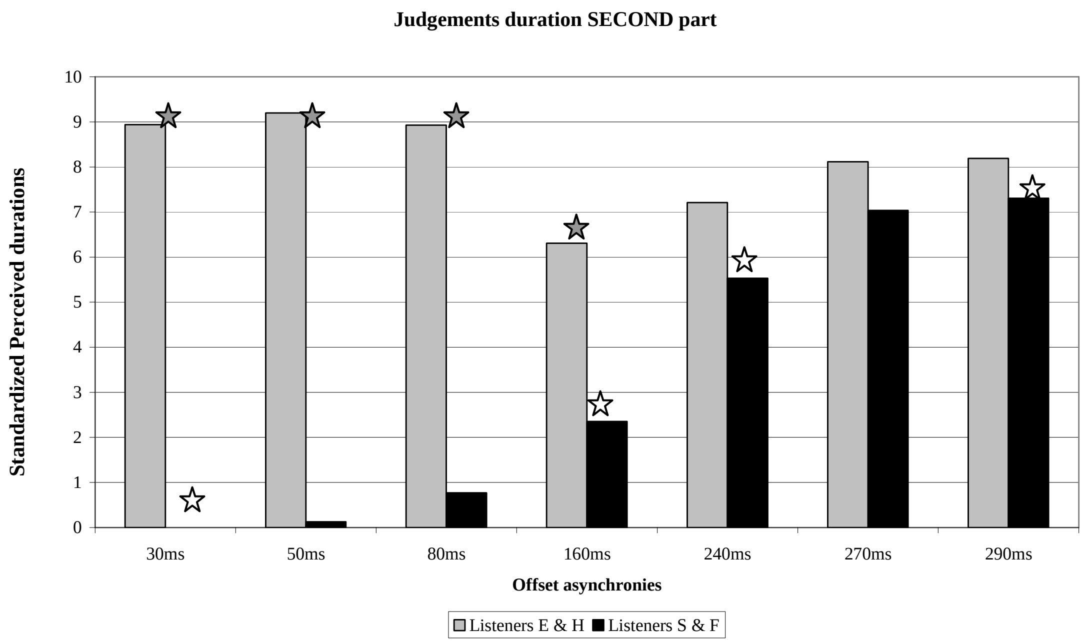

Fig. 5. Results for the judgements of the second part of the complex. Listeners E & H did not perceive the first three asynchronies: 30, 50, and 80 ms [F (1,4) = 5.27; F (1,12) = 7.21; F(1,5) = 12.32 respectively; all with a $p < 0.001$], therefore they judged the total duration of the complex tone. The offset asynchrony of 160 ms was overestimated [F (1,6) = 23.45, $p > 0.05$], while larger asynchronies showed the same tendency but did not reach statistical significance. Listeners S & H underestimated the duration of the 30, 160, 240 ms [F (1,14) = 4.67; F (1,12) = 6.46; F (1,23) = 56.87] respectively; all with a ($p > 0.01$) and 290 ms [F (1,2)=7.89, $p > 0.05$] asynchronies.

Overall, these results showed that the judgment of the second part of the complex, which should correspond to the stimulus part after the asynchrony, was less accurate and subjectively more difficult (listeners' personal report) than the judgement of the first part. Phenomenological data showed that listeners E & H could not detect asynchronies smaller than 160 ms; these data was further corroborated on the magnitude estimation task. The magnitude estimation task showed a task order effect and an overestimation tendency in the group performing condition 2 first. The underestimation observed in listeners S & F was probably due to an attentional bias to the fist part of the complex. The overestimation effect observed on the phenomenological data, and the tendency observed in the magnitude estimation task could be due to either the perceptual restoration of an onset or onset and filling (> 20 ms aprox.) or a critical minimum duration effect (Efron, 1970a). Efron (1970a) found that independently of the brevity of the stimulus duration, listeners always seem to perceive a critical length of 120-170 ms. That is, stimuli of less than 120-130 ms produced percepts which had durations identical to those produced by stimuli of 120-130 ms. More interestingly, in a second experiment (Efron, 1970b), when pairs of discontinuous stimuli of 500 ms of duration were presented sequentially, separated only by a small gap, listeners tend to perceive them as continuous if the gap was smaller than 50 ms. In this case, the gap inserted between stimuli was perceptually 'filled' with the offset of the preceding stimulus (perceptual offset latency). These results corroborate that the minimum duration of a gap, when the markers are identical, has to be longer than 40 ms to be perceived. Interestingly, the auditory system restored this 'ungrammatical' sequence by perceptually delaying the offset of the first stimuli by 40 ms. Efron concluded from a series of experiments that the perceptual offset delay is 40 ms longer than the onset delay, and that this duration should be included in the total critical stimulus duration (160-170 ms), that is, 120-130 ± 40 ms. These data reinforce the relationship between the matching temporal properties of duration of auditory events and temporal windows of integration [(TWI) Cowan, 1984]. Accordingly, our sensory percepts in audition do not correspond to the immediately present acoustic reality but rather to the outcome of temporal integration during a sliding window of 150-200 ms.

A new set of experiments are already in progress to solve the task order effect and clarify the overestimation tendency observed in listeners E & H.

Acknowledgement

This research was sponsored by JSPS (18 06800), in the fiscal years 2007 and 2008.

References

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

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, 96, 341-370.

Efron, R. (1970a) The relationship between the duration of a stimulus and a duration of a perception. *Neuropsychologica* 8, 37-55.

Efron, R. (1970b) Minimum duration of a percept. *Neuropsychologica*, 8, 57-63.

Handel, S. (1989). *Listening: An introduction to the Perception of Auditory Events.* London: MIT Press.

Koffka, K. (1935). Principles of Gestalt Psychology. Harcourt, Brace & World, New York.

Nakajima, Y. & Sasaki, T. (1995). A simple grammar of auditory streams. *Reports of the Acoustical Society of Japan*, H-95-75.

Saussure, F. de. (1966). *Course in General Linguistics*. NYC, USA: McGraw-Hill. (Original work published in 1916).