Title	The Underlying Frequency-extrapolation Mechanism on Auditory Processing
Author(s)	松井　理直（Michinao Matsui）
Citation	Theoretical and applied linguistics at Kobe Shoin, No.2：19-33
Issue Date	1999
Resource Type	Bulletin Paper / 紀要論文
Resource Version	
URL	
Right	
Additional Information	

The Underlying Frequency-extrapolation Mechanism on Auditory Processing

Michinao Matsui

The effect of auditory continuity was investigated using frequency glides preceding and following a noise burst. The results of the experiments suggested that subjects traced the frequency change of the prenoise glide and perceived its extrapolated trajectory during the noise. The maximum duration of the extrapolation was about 120 ms. The mechanism of frequency extrapolation may play an important role in the auditory scene analysis of a natural sound environment.

1. Introduction

In daily life we encounter various kinds of sounds, for instance speech sounds, musical sounds, traffic noise and so on. They seldom exist in isolation and are usually mixed with each other. To recognize speech sounds exactly in such a natural sound environment, it is necessary for our auditory system to extract the sound components we need and properly construct an auditory world. This function of the auditory system is known as auditory scene analysis (Bregman 1990), which consists of two auditory processes: they are primitive process and schema-based process.

One phenomenon that may reflect auditory scene analysis is auditory induction (Warren 1984). For example, when portions of a signal are obliterated completely by a louder noise, they can nevertheless, under appropriate conditions, be heard to be continuous during the noise. The same perceptual restoration can occur for speech signals: this is called phonemic restoration. In the latter case, linguistic knowledge that

Theoretical and Applied Linguistics at Kobe Shoin 2, 19–33, 1999.

belongs to schema-based process on auditory scene analysis has an important effect on the restoration process, in addition to primitive auditory process. Ray Jackendoff (1992) also places a special emphasis on the interaction between primitive process (auditory input) and schema-based process that involves both phonological structures as linguistic knowledge and grouping/metrical structures as musical knowledge.

Although auditory induction and phonemic restoration have been investigated under various conditions, their underlying mechanism has not been made clear enough. In this study, we focus on the auditory continuity of sinusoidal glides interrupted by a noise, where no linguistic knowledge can be used and peripheral responses such as masking are not sufficient to account for the continuity.

Furthermore, we tried to determine not only whether subjects judged the sequence to be continuous or not, but also what they perceived in the noise. To clarify such a dynamic and complicated phenomenon, however, the traditional methods of psychophysical measurement may not be appropriate. In this study, we introduced a method of drawing whereby subjects were required to express what they heard by drawing a line. Then, the frequency analysis mechanism suggested from the drawings was exiled, using traditional psychophysical methods.

2. Experiment 1

2.1 Method

The experimental design consisted of the following three factors: (1) initial frequency of the prenoise glide (Tone A), (2) initial frequency of the postnoise glide (Tone B), and (3) final frequency of the postnoise glide. The combination of these factors produced 18 stimuli in total (upward: 9 stimuli, downward 9 stimuli), as shown in Figure 1 (only downward).

The pre- and postnoise tones were sinusoidal glides with logarithmically changing frequency. Prenoise glide was either downward or upward. The initial frequency was 2000 or 500 Hz, and the final frequency just before the noise was always 1000 Hz. The duration of the prenoise glide was 500 ms, including a rise time of 20 ms.

The postnoise glide was upward, downward or steady-state. The initial frequency was 1515.7 Hz, 659.8 Hz or 1000 Hz, the first two values of which corresponded to the extrapolation of Tone A through the noise. The frequency range of the upward and downward glides was one octave, and the final frequencies are shown in Figure 1. The duration of Tone g was 500 ms, including a decay time of 20 ms.

The interrupting noise was a Gaussian noise lowpass-filtered at 5000 Hz with duration of 300 ms. The same "frozen" noise was repeatedly within each trial.

The sound pressure level of Tones A and B was 45 dB, and that of the noise was 65 dB. These stimuli were presented diotically to subjects in anechoic room via head-

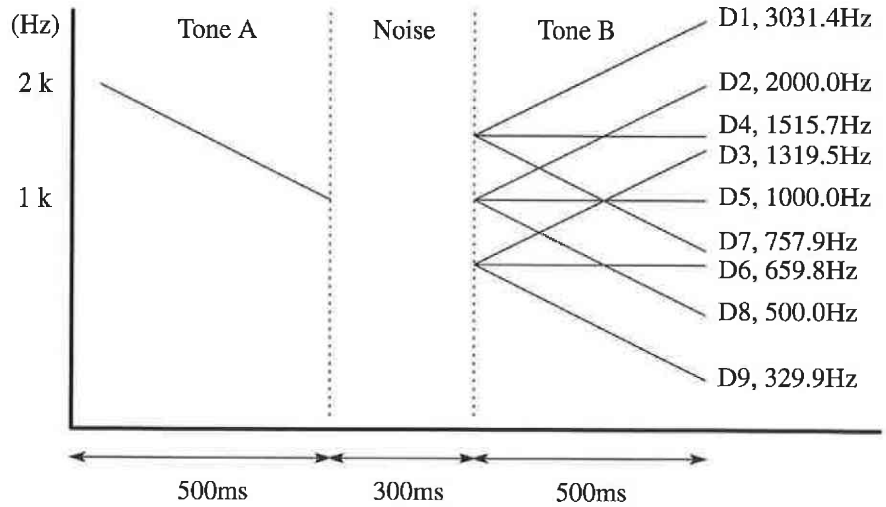

Figure 1: Schematic illustrations of the stimuli used in Experiment 1 (Tone A was downward). The name of each stimulus and its final frequency are displayed to the right of Tone B.

phones (STAX, SRD-X and SR-Λ pro).

The following method of drawing was introduced. Subjects were asked to express the trajectory they perceived during the noise and Tone B by drawing a line on a response sheet to create something similar to a sound spectrogram (Kurakata, Matsui and Nishimura 1997).

The trajectory of Tone A, the position of the noise, and the end of the tone had been drawn on the sheet in advance. Subjects used a mechanical pencil for drawing and were allowed to use an eraser. They were instructed to draw the trajectory as accurately as possible.

Each stimulus was presented repeatedly at 4-second intervals with each trial. The subjects were allowed to listen to the stimulus as many times as they liked. The 18 different stimuli were presented to each subject in random order. Twenty subjects with normal hearing ability participated in this experiment.

2.2 Results and discussion

Since the stimuli in panels (a) and (b) of Figure 1 are symmetrical to each other on the log-frequency axis and similar drawings were obtained for both stimulus groups, only the stimuli shown in panel (a), DI-9, are used in the analysis below.

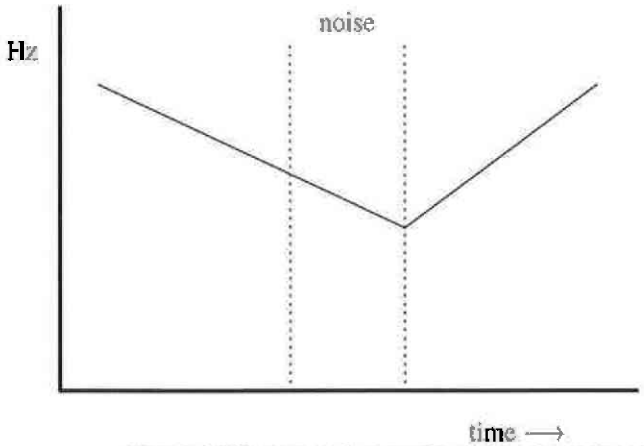

Figure 2: The typical example of drawings for the stimulus D2.

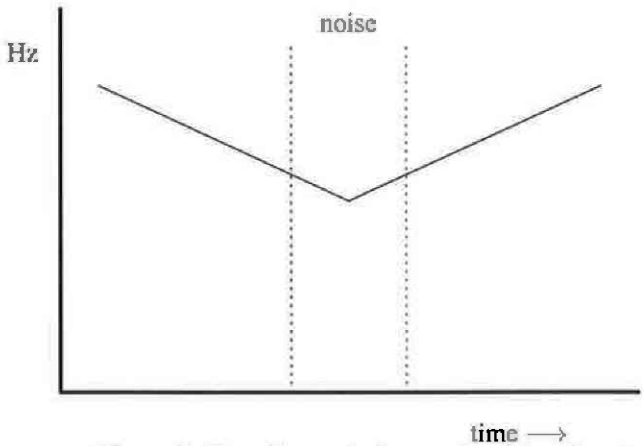

Figure 3: The other typical example of drawings for the stimulus D2.

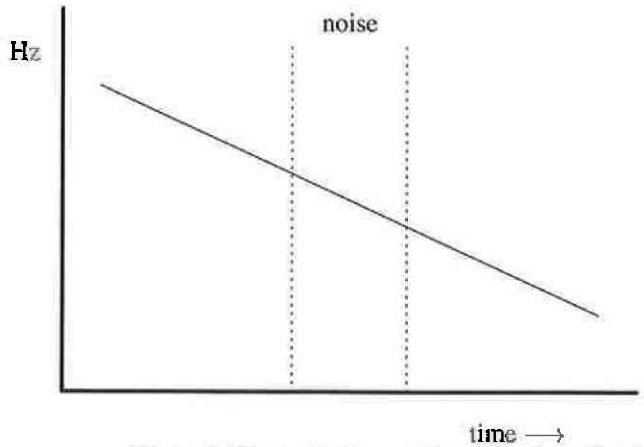

Figure 4: The typical example of drawings for the stimulus D8.

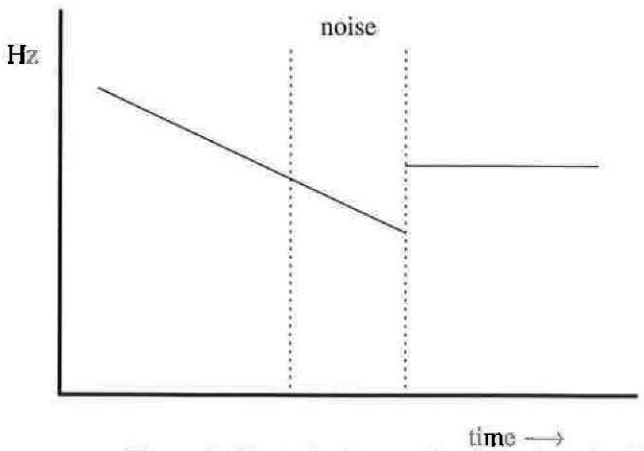

Figure 5: The typical example of drawings for the stimulus D5.

The stimuli were divided into three groups, according to their patterns of frequency change: (1) the direction of prenoise glide was different from that of postnoise glide [Dl-3]; (2) the direction of prenoise glide was the same as that of postnoise glide [D7-9]; and (3) postnoise glide was steady-state [D4-6].

In the group D1-3, two patterns of drawings appeared: (1) the trajectory of Tone A was extrapolated straight to the end of the noise (Figure 2), and (2) the trajectory of Tone A was extrapolated, but turned upward in the middle of the noise (Figure 3). Each pattern accounted for about 40 percent of drawings in this group. These drawings seem to show that the subjects did not simply interpolate the two points entering and exiting from the noise, as Ciocca and Bregman (1987) suggest, but extrapolated, to some extent, the preceding glide through the noise.

In the group D7-9, about 90 percent of the drawings showed that the trajectory of Tone A went straight or wound a little through the noise (Figure 4). In particular, almost subjects drew a straight trajectory for the stimulus D9, suggesting a strong continuity.

In the group D4-6, two thirds of the drawings showed that the trajectory of Tone A was extrapolated to the end of the noise, and then a new trajectory started for Tone B (Figure 5). Although there were other patterns of trajectory, most of them did not show any pause in the noise, and the subjects seemed to perceive the tone to be present.

These patterns of trajectory can be explained by hypothesizing the following three processes: (1) Based on the frequency change of Tone A, an extrapolated trajectory in the noise is determined temporarily; then (2) this tentative trajectory is adjusted to connect with Tone B smoothly (Figures 2, 3 and 4); (3) If the gap between the final pitch of the trajectory and the initial pitch of tone B is too large, the tentatively determined trajectory is adopted as final without being affected by Tone B (Figure 5).

Although the hypothesis of a frequency-extrapolation mechanism may explain the drawings quite well, it is not clear whether or not the subjects really perceived the extrapolated trajectory that was not physically present in the noise.

Moreover, the method of drawing is convenient and efficient for the investigation of the perception of these complicated sounds, but the results are heavily dependent on the drawing ability of the subjects. This problem may knit the reliability of the interpretation of the drawings.

To examine the validity of the frequency-extrapolation mechanism and the degree of its effect, the following two psychophysical experiments were conducted.

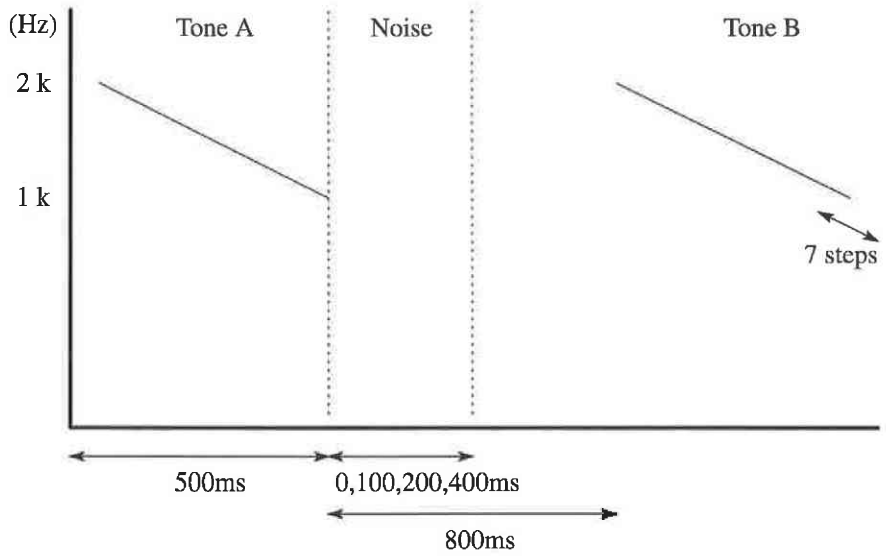

Figure 6: Schematic illustration of the stimuli used in Experiment 2-1. Tone B was varied in seven steps.

3. Experiment 2-1

3.1 Method

The stimuli used are shown in Figure 3. The standard stimulus, Tone A, was a downward glide moving from 2000 Hz to 1000 Hz. Its duration was 500 ms, including a rise time of 20 ms. Tone A was followed by a lowpass-filtered Gaussian noise with a cut-off frequency of 5800 Hz. The duration of the noise was 0 ms (without noise), 100 ms, 200 ms or 400 ms. The same "frozen" noise was used repeatedly within each trial. The comparison stimulus, Tone B, was a downward glide with the same slope as Tone A. Its duration was varied in seven steps centered around 500 ms.

The method of constant stimuli was used. One of the seven series of duration of Tone B was randomly presented to subjects. After listening to the whole sequence, the subjects were required to indicate whether the duration of Tone B was "longer" or "shorter" than that of Tone A. Every subject performed 20 trials for each comparison stimulus. Eight subjects with normal hearing ability participated in this experiment.

3.2 Results and discussion

The PSE of each subject for every standard stimulus was calculated by the method of maximum likelihood. Then, the difference between the PSE of without noise condition

and that of each duration of noise was defined as the amount of overestimation. The overestimation and standard deviations among subjects as a function of noise duration are shown in Figure 4. It can be seen that the dura6on of Tone A was overestimated for all conditions, which suggests that Tone A was perceptually extended into the noise. The amount of overestimation inquired as the duration of the noise increased.

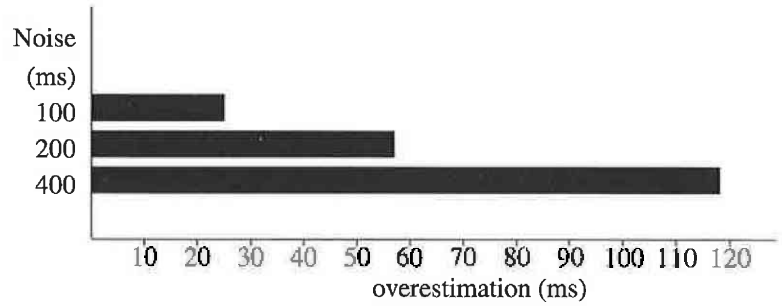

Figure 7: Overestimations and standard deviations of the duration of Tone A as a function of noise duration.

This result, however, may not necessarily confirm the existence of an extrapolated trajectory. A certain delay in response to Tone A could occur at some stage of auditory processing, and the overestimation might simply reflect this "after-effect." If the trajectory of Tone A is ray extrapolated, the final pitch of Tone A must be perceived to be lower than it actually is. To clay this point, another experiment was conducted.

4. Experiment 2-2

4.1 Method

The stimuli used are shown in Figure 8. They were essentially identical to those used in former experiment, except for the comparison stimulus, Tone B. Tone B was a steady-state tone sting 800 ms after the end of Tone A. Its frequency was varied in seven steps centered around 1000 Hz.

The method of constant stimuli was used. One of the seven frequencies of Tone B was randomly presented to the subjects. After listening to the whole sequence, the subjects were required to indicate whether the pitch of Tone B was "higher" or "lower" than the final pitch of Tone A. Every subject performed 20 trials for each comparison stimulus. Eight subjects with normal hearing ability participated in this experiment.

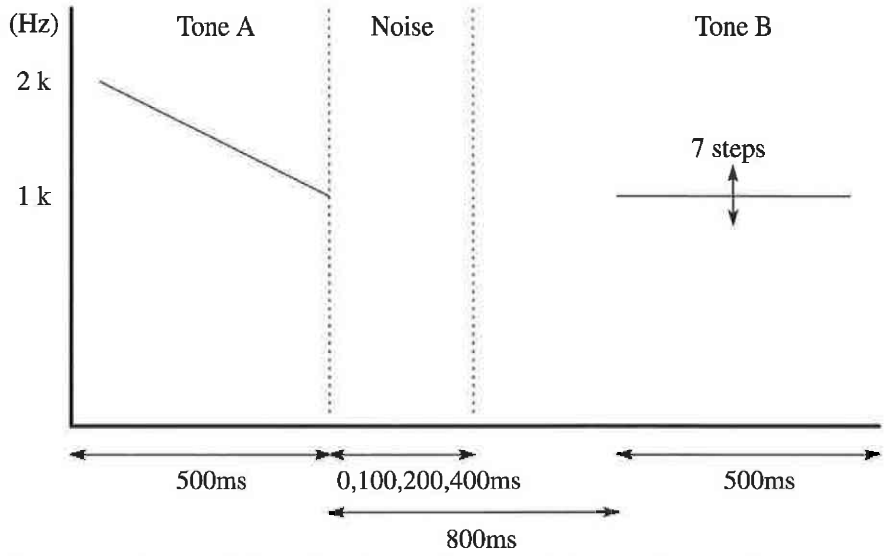

Figure 8: Schematic illustration of the stimuli used in Experiment 2-2. Tone B was varied in seven steps.

4.2 Results and discussion

The PSE of each subject for every standard stimulus was calculated by the method of maximum likelihood. The PSEs and standard deviations among subjects as a function of noise duration are shown in Figure 9. It can be seen that the final pitch of Tone A was perceived to be lower than it actually was. The discrepancy between the PSE and actual frequency tended to become larger as the duration of the noise increased.

For comparison with the results of former experiment, the PSE of pitch was converted to the amount of overestimation of duration, as shown in Figure 10. This graph shows the same tendency for overestimation that was seen in former experiment, although the values in this experiment were a little larger. Therefore, we may conclude that the subjects did perceive an extrapolated trajectory in the noise, and did not simply have a delayed response to Tone A.

4.3 A factor of phonemic restoration

Even when portions of speech sounds, as well as auditory induction of a glide tone, are obliterated completely by a louder noise, they can nevertheless be heard to be continuous during the noise. This phenomenon called phonemic restoration is affected by not only linguistic knowledge but also primitive auditory process. Masuda, Aikawa

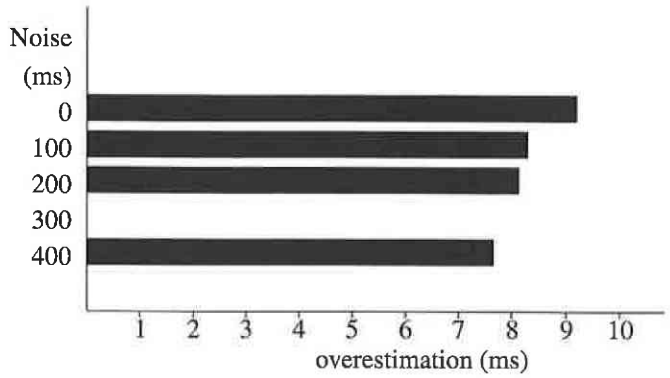

Figure 9: PSEs and Standard deviations of the final pitch of Tone A as a function of noise duration.

and Tsuzaki (1995) shows that following noises inflect the perception of phonemic categories at the transient position in countinuous speech sounds. It is reasonable to suppose that the mechanism of perceptual extrapolation we have discussed in the preceding sections causes this perceptual changes of phonemes. Their experiment is an appropriate sample of phonemic restoration affected by auditory primitive processing.

5. A Cognitive Model of Auditory Scene Analysis

5.1 Features of an Auditory Processing Model

In this section, we sketch an outline of constraints and a simple cognitive model of auditory scene analysis. The analysis seems to consist of two stages, a primitive process and a schema-based process (Bregman 1990). Although these two stages can not be distinctly separable and may interact with each other. Now, we want to illustrate (1) what acoustical cues for the analysis are used and (2) how these cues interact with each other in the analysis over time.

5.2 Constraint-based Model

Hashida and Matsubara (1994) argued that any system which deals with a complex and "ill-formed problem" such as partiality of information must be designed based on the constraints. "Constraints" they named have the following characteristics: (1) A constraint is orthogonal to the flow of information so that it can accommodate a non-modular flow of information. (2) A constraint describes both the inside and outside of the system, because it refers to neither input nor output. (3) A constraint corresponds

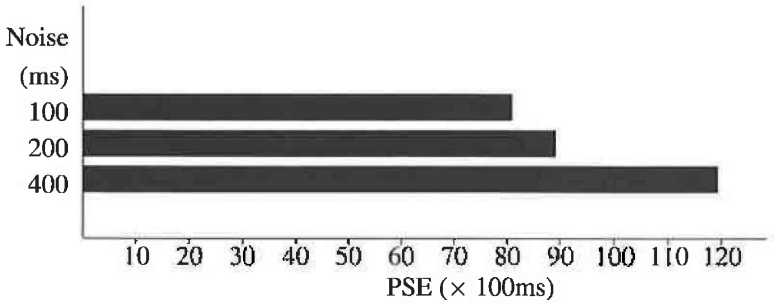

Figure 10: Overestimations and standard deviations of the duration of Tone A as a function of noise duration.

to feedback loops connecting the cognitive system and its environment.

This principle of design enables a system to adapt all sorts of general conditions. Even under novel situations, a system would give some valid solution without being stopped completely. For natural sound environments, however, all the constraints are not always satisfied. Therefore, we propose an auditory processing model that allows violations of the constraints.

This model has the following stages: (1) At a lower level of auditory processing, a set of candidates which is to be the output of a given input signal. (2) The relative well-formedness of a cognitive object is measured by the "harmony" of the object. The cognitive system can therefore be specified by the harmony-function itself, which is equivalent to a harmony-maximizing connectionist network (Prince and Smolensky 1993). (3) Every constraint determining the well-formedness is given a certain position in a ranking from low to high. (4) The constraints allow violations. The degree of the violation, however, must be minimized. The degree is determined according to the ranking of the constraint; the higher the violated constraint, the more serious the violation is. (5) The degree of violation is calculated for every solution candidate. The candidate which has the least violation is selected.

5.3 Constraints on the Stream Forming Process

The primitive process of auditory scene analysis consists of two organizations, simultaneous stream organization and sequential stream organization. A "stream" here means a perceptual unit that represents a single auditory event. In this model, we assume that these two modules work in parallel and interact with each other to make a final auditory image.

Let us first consider the constraints in the simultaneous stream organization module. The duration of the window of this module has been estimated to be about 100 msec (Matsui et al., 1996). This simultaneous stream organization module outputs a set of solution candidates ranked according to the level of violation of the constants. This process has the following constraints and ranking, whose relative significance among constraints is indicated by the inequality sign, "≫".

Onset/offset synchrony ≫
solution of the preceding sequential streaming ≫
common AM or FM harmonicity ≫
direction of frequency change ≫
number of formed streams ≫ ....

These constraints are defined as follows.

- Onset/offset synchrony: Components which start or stop at the same time tend to form a stream.
- Solution of the preceding sequential streaming: The preceding time window offers a possible solution to the current window.
- Common AM or FM: Components which have the same amplitude-modulation (AM) or frequency-modulation (FM) tend to form a stream.
- Harmonicity: Components which constitute a harmonic structure tend to form a stream.
- Direction of frequency change: Components whose frequencies change in the same direction tend to form a stream.
- Number of formed streams: The number of formed streams at one time should be as few as possible.

On the other hand, the ranking of constraints in the sequential stream organization module is much more complicated than that in the simultaneous module. In this module, it cannot be assumed that there is a fixed ranking common to everyone, because schemata such as linguistic or musical knowledge can be deeply involved. Therefore, we concentrate here on some of the innate constraints, not on those dependent on acquired knowledge.

The time window of this module has been estimated to be about 400 msec (Kurakata et at., 1994, 1995), during this time period, possible candidates are successively

stored. Then, after calculating which candidate is the most suitable, the module outputs it as a final (that is, a finally determined auditory image). The constraints ranked in sequential stream organization could be as follows:

Similarity of frequency range $\gg$
proximity of time $=$
proximity of frequency $\gg$
ranking of the candidates of simultaneous streaming $\gg$ ....

These constraints can be defined as follows.

- Similarity of frequency range: Components which occupy the same frequency range am captured to form a stream.
- Proximity of time: Components which are close to one another in time tend to form a stream.
- Proximity of frequency: Components which are close to one another in frequency tend to form a stream.
- Ranking of the candidates of simultaneous streaming: The ranked solutions offered by the simultaneous module of the current time window affect the streaming.

We should notice that the solution of the preceding sequential streaming itself functions as a constraint of simultaneous segregation and that the ranking of the candidates of simultaneous segregation is directly incorporated as a constraint of sequential streaming. This relationship shows that the solution of the preceding steam is a dominant constraint, a candidate which has a good context-dependency can be ranked at a high position. Furthermore, this enables both the simultaneous and sequential stream organization to be involved in the constraints of the model and to interactively cope with the problem of auditory scene analysis. This interaction between simultaneous segregation and sequential streaming is the most important cause of auditory continuity shown in former sections.

6. General Discussion

It has been doubted that a frequency-extrapolation mechanism is involved in the process of auditory continuity (e.g., Dannenbring, 1976). However, there are two significant differences in experimental design between previous studies showing negative results and the present study. First, in previous studies, repeating cycles were often adopted as stimuli; however we presented each stimulus separately, at a time interval,

because of the possibility that the perception; of continuity could change during the repetition. Second, previous studies used a noise preceded and followed by a tone, whereas we deliberately removed the postnoise tone in the last two experiments in order to avoid backward influence of that tone on the perceptual trajectory during the noise. These conditions in the previous studies may have worked against the frequency extrapolation mechanism.

Auditory continuity has generally been explained by the postulation that the trajectory underneath a noise is determined "in a retrograde fashion" (Ciocca and Bregman, 1987), after listening to the whole sequence. From an ecological point of view, however, this would be impractical in cases where the process for a target sound is completely interrupted by an intrusion of loud external noise. Estimating the following trajectory of sound based on the preceding information would help our auditory system cope with the sound mixture in a natural environment. This mechanism may underlie perceptual continuity in general, as well as that of gliding tones in particular, and is included in the cognitive model of auditory scene analysis.

The purpose of this cognitive model is to give an overview of auditory scene analysis whose aspects have been investigated separately. However, there are many other possible constraints that can affect streaming, although we have not incorporated them into our current model. Among them are similarity of loudness, similarity of timbre, location of sound source, and the rhythmic structure of a stream. The reasons why we have not used them in the model are that the degree of their effect is not clear enough to determine their position in the rank, and because terms such as "timbre" and "rhythm" themselves are ambiguous compared to other factors. Further psychoacoustical studies will be required before we can incorporate these factors into the model of auditory scene analysis.

References

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. in S. McAdams and E. Bigand (eds). *Thinking in Sound*, pp.10–36. New York, N.Y.: Oxford University Press.

Ciocca, V. and Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception and Psychophysics.*, 42, 476–484.

Cooke, M. (1993). *Modelling Auditory Processing and Organisation.* Cambridge: Cambridge University Press.

Dannenbring, G. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology.*, 30, 99-114.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of Acoustical society of America.*, 76, 1636–1647.

Hasida, K. and Matsubara, H. (1994). An essay on the design principle of intelligence: Partiality, constraint, and the frame problem advances. *Japanese Cognitive Science.*, vol.7., 159–201. Tokyo: Kodansha Scientifics. (in Japanese).

Jackendoff, R. (1992). *Languages of the Mind: Essays on Mental Representation.* MIT Press.

Kurakata, K., Matsui, M. and Nishimura, A. (1997). *Proceedings of first international conference on cognitive science.* pp.183–187. Seoul University.

Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of Acoustic Society of America*, 22 (6), 824–826.

Liberman, A. M. and Cooper, F. S. (1962). A motor theory of speech perception. *Proceedings of Speech Communication Seminar*, paper-D3, Stockholm.

Masuda, I., Aikawa, K. and Tsuzaki, M. (1995). Effect of fllowing noise on the perception of transient in continuous speech sounds. *Proc. Autumun Meet. Acoustic Society of Japan.* (in Japanese).

Matsui, M., Kurakata, K. and Nishimura, A. (1996). Some constraints on a cognitive model of auditory scene analysis. *Proc. Spring Meet. Acoustic Society of Japan.*, 485–486. (in Japanese).

Matui, M. F. (1997). An Overview to JPSG Phonology. In Gunji and Hasida (eds.) *Topics in constraint-based grammar of Japanese.* Kluwer Academic Publishers. pp. 99–140.

Prince, A. and P. Smolensky. (1993). *Optimality Theory : Constraint Interaction in Generative Grammar.* Technical report 2, Center for Cognitive Science, Rutgers University.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. Psychological Bulletin., 96, 371–383.