AUDITORY GRAMMAR

Yoshitaka Nakajima[1], Takayuki Sasaki[2], Kazuo Ueda[1], and Gerard B. Remijn[1]

[1]Department of Human Science/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka 815-8540, Japan

[2]Department of Psychological and Behavioral Science, Miyagi Gakuin Women's University, Sendai 981-8557, Japan

nakajima@design.kyushu-u.ac.jp

Auditory streams are considered basic units of auditory percepts, and an auditory stream is a concatenation of auditory events and silences. In our recent book, we proposed a theoretical framework in which auditory units equal to or smaller than auditory events, i.e., auditory subevents, are integrated linearly to form auditory streams. A simple grammar, Auditory Grammar, was introduced to avoid nonsense chains of subevents, e.g., a silence succeeded immediately by an offset (a termination); a silence represents a state without a sound, and to put an offset, i.e., the end of a sound, immediately after that should be prohibited as ungrammatical. By assuming a few gestalt principles including the proximity principle and this grammar, we are able to interpret or reinterpret some auditory phenomena from a unified viewpoint, such as the gap transfer illusion, the split-off phenomenon, the auditory continuity effect, and perceptual extraction of a melody in a very reverberant room.

INTRODUCTION

If we try to record and write down what someone says in an everyday conversation, we are very likely to be embarrassed by the fact that the speech includes a fair amount of doubtful parts in terms of grammar. Probably, in our mind we often correct what we hear in an everyday situation according to a grammatical framework, which needs to be shared with the social group to which we belong. Because acoustic information disappears immediately after it is released, often in a noisy environment, the auditory system simply needs a robust framework to connect given pieces of information in a coherent manner. If so, however, the auditory system may need such a framework in order to organize any auditory percept in our everyday life—for example, to hear out footsteps, approaching cars, cats' meows, winds, sound signals of electronic devices, and so on. It is indeed an astonishing capacity of the human auditory system to separate each sound perceptually when mixtures of many different sounds are given to both ears simply as temporal changes of sound pressure [1].

Our research over the past years proceeded from the hypothesis that our auditory system utilizes a kind of grammatical system which is innate to humans, and that this system is a basis of all specific grammars of human languages. So far, the hypothesis is still very primitive, but it helped us to understand and discover new auditory illusions. It seems to have a path to be connected to the phonologies of English, Japanese, or Chinese, and seems to explain partially how notes in Western music are perceived. Some neurophysiological phenomena can be related to this human innate grammar. Thus we called this the *Auditory Grammar*, abbreviated as AG from here on, and wrote a book in Japanese to sum up what is known in relation to this paradigm [2]. An outline of this book is described in this article.

The concept of gestalt quality appeared at the end of the 19th century to explain the fact that one can perceive the same melody in different keys, e.g., in C major and in F# major, even if no common notes are used in two different presentations, e.g., "C D E C | C D E C" and "F#G#A#F# | F#G#A#F#" [3] (Figure 1). Something that cannot be reduced to the natures of individual tones should be there, and this was called the gestalt quality. This was an immediate precursor of gestalt psychology, which appeared as a rather quiet scientific revolution claiming that the whole is not the sum of its parts (e.g., Koffka [4]). The contemporary leading researchers in auditory psychology were not interested in this idea, and rather established a theoretical framework in which auditory phenomena were interpreted as if they had been phenomena observed in an electric circuit [5]. To be fair, this paradigm worked very well for decades to make auditory research a very rigorous and precise field [6].

However, a few related fields could not afford neglecting gestalt psychology. In the field of speech perception, two phenomena, i.e., the cocktail party effect [7] and the auditory continuity effect [8] were reported. The former is a common phenomenon in our everyday life. When two or more people speak different things simultaneously, we are able to perceive that more than one speaker utters different things, and follow one of the speakers to grasp the spoken content. The latter is now a well-known auditory illusion: a speech or music signal, a tone, or a band noise of which a short portion, typically a small fraction of a second, is replaced with an intervening noise can be perceived as continuous, although that portion is missing. In order for this illusion to occur, the noise to fill the missing part should basically cover the frequency range of the original signal with a surpassing intensity. Another important phenomenon related to gestalt psychology has been reported with some interest in music [9]. If two pure tones of 100 ms alternate between 1000 and 1050 Hz, we are likely to hear a single pitch-fluctuating tone as if we hear a trill in music. If the tone frequencies are 1000 and 2100 Hz instead, we are likely to hear two separate streams of different pitches. The latter phenomenon is called *auditory stream segregation* today [1]. Auditory stream segregation is often understood employing a gestalt concept called the *proximity principle*: objects or events that are close to each other tend to be integrated

perceptually. Deutsch [10] systematically indicated that those gestalt principles established to understand visual organization in the first half of the 20th century, to which the *similarity principle* and the *common fate principle* are also included, work well to understand auditory organisation especially in music perception.

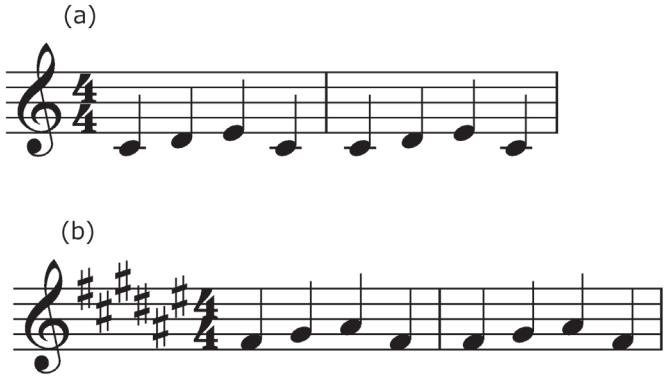

Figure 1. An example showing the concept of gestalt quality [3]. Even the transposition to the remotest key does not prevent the listener from hearing the same melody, although none of the notes are shared between the two tone sequences.

AUDITORY UNITS

In visual perception, figures and a ground are often formed perceptually to let our visual world make sense. For example, the letters on this page are figures, and they are supported by a ground throughout the page, which includes the parts covered by the letters. The ground does not have a clear shape. The figure-ground idea is often applied to auditory organization—for example, a melody and an accompaniment are sometimes considered as a figure and a ground. We do not take this view because both the melody and the accompaniment have clear shapes. Basically, no parts of the accompaniment are covered and hidden by the melody, and we can even pay attention only to the accompaniment for a long time. We rather assume that the auditory world consists of auditory streams that are concatenations of auditory events and silences. This is not a revolutionary way of thinking, but we just formalized what leading researchers assumed on auditory organization [1, 11, 12]. Auditory events are what we often call sounds in our everyday life: footsteps, hand claps, music notes, or speech syllables, of which we can count the number. An auditory stream is what we hear coherently, often as belonging to the same source, in time, which is a string of auditory events and silences. Auditory events and auditory streams are perceptual units comprising the auditory world.

In order to formalize AG, we took one further step to assume auditory elements that can be smaller than auditory events, called *auditory subevents*. AG describes how such elements are concatenated to form auditory events and streams. Besides gestalt principles, our ideas about auditory subevents seem to have taken shape along with approaches in neuroscientific research on how the human brain deals with incoming sound. It has long been described in various neuroscientific studies that sound edges, i.e. a sound's onset and offset, are signaled at very early stages of cortical processing by cells that only respond to sound edges. These edge-specific neurons are typically different from those that signal the sustained parts of sound, indicating that the brain considers sound edges and sustained parts as different subevents. Current research addresses not only how sound edges are represented in the brain, but also how the information of sound edges and the information of sustained parts of sound are combined and expressed at the level of cortical responses over time. For example, the auditory continuity effect as introduced above and described in more detail below, has been studied specifically to investigate such neural-response integration [13, 14]. Conceptually, AG also proceeds from the integration of sound parts that together constitute auditory events and auditory streams. This enabled us to study the integration of auditory subevents at the behavioral (psychophysical) level by means of creating new auditory phenomena, such as those described below. In the future, similar to the auditory continuity effect, these new sound stimuli can hopefully be subjected to and contribute to neuroscientific research as well.

GAP TRANSFER ILLUSION

An auditory illusion was the starting point to construct AG (Figure 2). Suppose that a frequency glide component of 2500 ms moving from 420.4 to 2378.4 Hz and another glide component of 500 ms moving from 1189.2 to 840.9 Hz cross each other while sharing their temporal middle. This pattern is typically perceived as a long ascending glide and a short descending glide crossing each other—just as how the stimulus pattern was made. These glides physically cross at 1000 Hz, but "crossing" in the present context means that there are ascending and descending glides which share the same pitch region. If a short temporal gap of about 100 ms was introduced onto the middle of the short glide, we hear what we presented—a long ascending glide and two successive short tones. However, if a short temporal gap was introduced onto the middle of the long instead of the short glide, we still hear a long continuously ascending glide and two successive short tones. This is the *gap transfer illusion*, which gave us a chance to think about AG [15].

The long glide with a temporal gap is described by the following scheme:

```
abcdefghijklmnop
<  =  > <  =  >/,
```

where the alphabetic letters indicate temporal positions roughly, and "<" means an onset, "=" a filling, ">" an offset (a termination), and "/" a silence. The short glide is added as follows:

```
abcdefghijklmnop
<  =  > <  =  >/
    <  =  >/    .
```

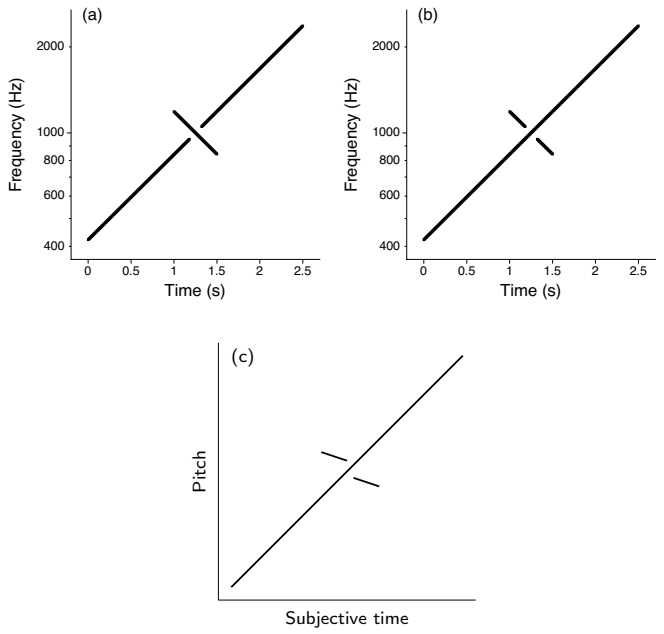

Figure 2. The gap transfer illusion. (a) A typical stimulus pattern inducing the illusion, (b) a stimulus pattern perceived as it is, and (c) the common percept.

Because the onset of the short glide (at the letter "e") and the offset of the first part of the long glide (at "g") are close to each other in time and frequency, the proximity principle works to integrate them perceptually to make an auditory event. The onset of the second part of the long glide (at "i") and the offset of the short glide (at "k") are also close to each other, and again the proximity principle works to integrate them. Thus, we obtain:

```
abcdefghijklmnop
    <=> <=>     ,
```

and two auditory streams are formed as follows:

```
abcdefghijklmnop
<       =       >/
    <=>/<=>/    .
```

It is also possible that the silences (/) in the second line are detected at the beginning, but this does not change the final results. The potentially separate pieces of fillings are integrated as a single filling in the above stream, and this also helps to make the whole pattern grammatical. This shows how a grammatical form of an auditory percept appears from acoustic cues, which are not always grammatical.

AUDITORY GRAMMAR

AG is a grammar indicating how auditory subevents, i.e., onsets (<), offsets (>), fillings (=), and silences (/), are concatenated to form an auditory stream. We first assumed that an auditory stream always begins with an onset, and ends with a silence, and, then formalized a grammar as follows:

1. An onset is followed by a filling or a silence (<= or </).
2. An offset is followed by a silence (>/).
3. A filling is followed by an offset or an onset (=> or =<).
4. A silence is followed by an onset, or ends a stream (/< or /).

This set of rules may be insufficient for future research, but we first summarized what is known empirically, and did not include unnecessary rules for this purpose. New rules may be included in the future.

Three different types of auditory events appear as follows:

1. An event that begins and ends immediately as a single hand clap (<)—followed by a silence (/).
2. An event that begins, continues for a while, and ends as a train whistle (<=>)—followed by a silence (/).
3. An event that begins, continues for a while (<=), and is replaced by another event—starting with an onset (<)—as a music note in a melody.

The gap transfer illusion as described above is considered an auditory phenomenon to construct auditory events of the second type (<=>). If a filling and an offset appear *without* a preceding onset (=>), then AG requires an onset to be restored (<=>; Sasaki et al. [16]).

SPLIT-OFF PHENOMENON

A new illusion was discovered from this theoretical framework. Suppose that a long glide of 1200 ms moves from 420.4 to 965.9 Hz—the first part of the above long glide interrupted by a gap. The beginning part of the second glide is the same as that of the above short glide, but the glide is lengthened—another glide of 1500 ms moves from 1189.2 to 420.4 Hz. These two glides are presented successively with an overlap of 200 ms (Figure 3). This is a new example for this article, but the basic idea was from our previous research [15, 17]. We can hear a long continuous glide going up and down. At about the temporal middle of this ascending-descending glide, we hear a short tone. The acoustic cues of this pattern are:

```
abcdefghijkl
<  =  >
    <  =  >/.
```

The proximity principle works between the onset of the second glide (at "e") and the offset of the first glide (at "g"), and the following streams are obtained:

```
abcdefghijkl
<    =    >/
    <=>/    .
```

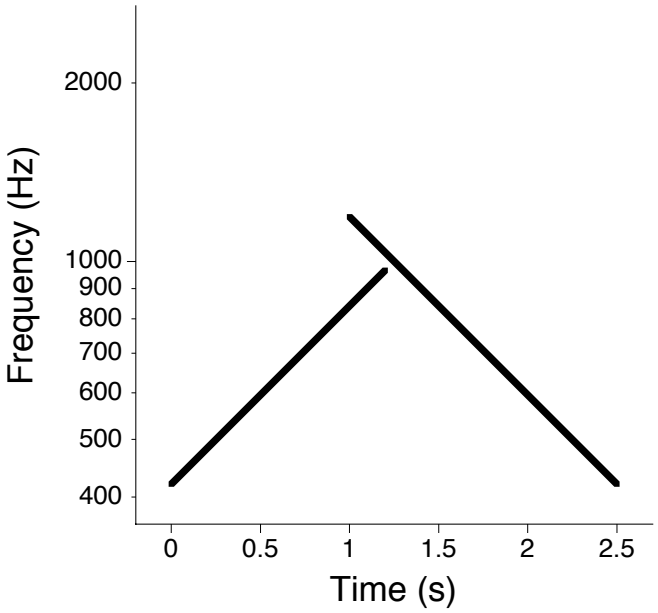

Figure 3. A stimulus pattern inducing the split-off phenomenon. An illusory short tone is perceived around the temporal middle of the pattern.

This is what we call the *split-off phenomenon*. Although we invoked one of the gestalt principles, the proximity principle, our explanation of this phenomenon also revealed a problem of gestalt psychology. The acoustic cues as indicated at first seem to take a very simple shape—two glides with a short overlap. It is difficult from a gestalt-psychological viewpoint to understand why the perceptual system should reconstruct this configuration. There is no reason to cause the split-off phenomenon in order to meet the *Prägnanz* law, which indicates a tendency of our perceptual system to seek for simplicity and regularity [4]. This led us to assume that the proximity principle and AG should be the basis to understand this phenomenon. The proximity between auditory subevents may have high priority in the process of auditory organization, and the proximity principle should not be chained to the classic version of gestalt psychology. AG, though still primitive, is justified by the fact that it gives us opportunities to discover new auditory phenomena. The split-off phenomenon can be observed in a very simple situation which could have been realized in the middle of the 20th century, but it seems that previous researchers did not have an occasion to generate a pattern leading to this illusion.

AUDITORY CONTINUITY EFFECT

If a long pure tone of 2000 ms and 2000 Hz has a temporal gap of 100 ms in the middle, and if the gap is filled with a narrow-band noise around 2000 Hz sufficiently more intense than the pure tone, then we often hear the pure tone not with a temporal gap but as continuous (Figure 4). This is an example of the *auditory continuity effect* [1, 8, 12]. This illusion is often explained by the peripheral behavior of the auditory system, but it seems worthwhile to indicate that this illusion can also be explained within our framework—AG should be always counted as one of the possible explanations.

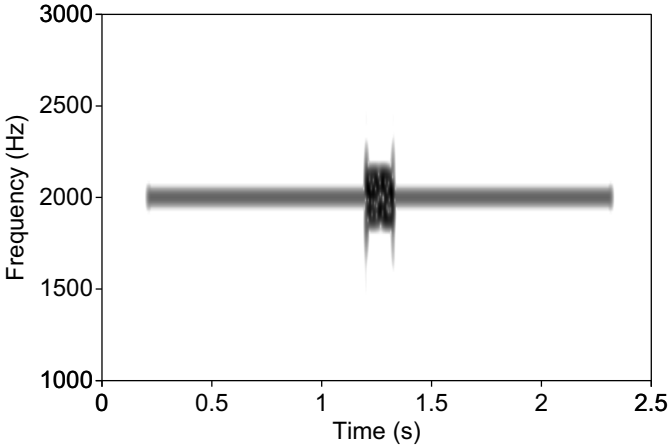

Figure 4. A spectrogram of a stimulus pattern in which the auditory continuity effect can be observed. A long continuous pure tone is perceived, even when the middle portion of 100 ms is replaced by a narrow-band noise of the same duration.

Because the intense noise should mask the offset and the onset of the pure tone portions delimiting the temporal gap, the following subevent cues are given to the auditory system:

```
abcdefghijkl
< = <=> = >/.
```

This configuration is ungrammatical. The first offset (at "g") should be followed by a silence (/), but is not. Then the configuration can be reconstructed, and a silence can be inserted (at "h") as follows:

```
abcdefghijkl
< =     = >/
    <=>/    .
```

This is closer to a grammatical solution. Because the two fillings, or filling portions, in the upper stream (at "c" and "i") are from the same pure tone, they can be united as a single filling. Thus, we obtain:

```
abcdefghijkl
<    =    >/
    <=>/    .
```

This shows that the auditory continuity effect can be understood also in the framework of AG [17–19].

PERCEPTUAL EXTRACTION OF A MELODY

Finally, we would like to point out an auditory phenomenon, which we should be encountering often in our everyday life, especially in music. Suppose that a pure tone of 1047 Hz (C in music) and 1200 ms and another pure tone of 988 Hz (B in music) and 1000 ms are presented in such a way that they share an offset (Figure 5). We can hear two tones with different pitches, just as how the tones are presented, with asynchronous onsets:

```
abcdefghij
<    =  >
        /
  <  =  > .
```

However, it is also possible to hear out a sequence of two successive notes C and B:

```
abcdefghij
<=<  =  >/
C-B------
   (roughness).
```

The auditory system probably tries to find a coherent stream to make the percept as simple as possible. B is perceived with some roughness, which means that C (to begin at "a") and B (to begin at "c") have a perceptual interaction, but the presence of C is suppressed perceptually when B starts. A very simple melody "CB—" thus appears. If we played this melody "CB—" in an extremely reverberant room with a recorder, for example, a very similar acoustic pattern would appear, and to perceive a melody in this way should be *correct* in this case.

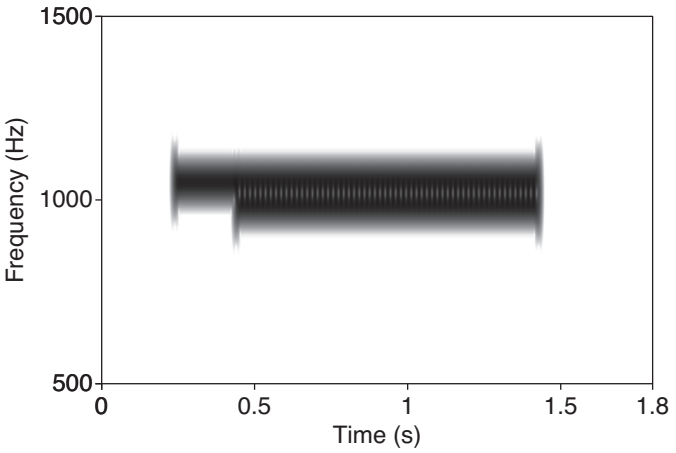

Figure 5. A subjectively constructed melody. Two pure tones of 1047 Hz (C in music) and 988 Hz (B in music) are presented with a 200-ms onset asynchrony. It is possible to hear a melody of C followed by B.

FINAL REMARKS

Some gestalt principles and AG can work together well to discover new auditory phenomena and to reinterpret our auditory experience. We invite people in many different fields to play with these toys to find something new themselves.

ACKNOWLEDGMENTS

Gert ten Hoopen worked with us to develop this theoretical framework. We are grateful to Kate Stevens for this opportunity to present part of our ideas in English. This study is supported by a Grant-in-Aid (25242002 to YN in FYs 2013-2017) from the Japan Society for the Promotion of Science.

REFERENCES

[1] A.S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambridge, MA, 1990

[2] Y. Nakajima, T. Sasaki, K. Ueda, and G.B. Remijn, *Auditory Grammar*, Corona Publishing, Tokyo, 2014 (in Japanese, with a CD-ROM)

[3] C.v. Ehrenfels, "Über 'Gestaltqualitäten'", *Vierteljahrsschrift für wissenshaftliche Philosophie*, **14**, 242-92 (1890)

[4] K. Koffka, *Principles of Gestalt Psychology*, Routledge and Kegan Paul, London, 1935

[5] S.S. Stevens and H. Davis, *Hearing: Its Psychology and Physiology*, Wiley and Sons, New York, 1938

[6] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd edition, Springer, Berlin, 2007

[7] E.C. Cherry, "Some experiments on the recognition of speech, with one and with two ears", *Journal of the Acoustical Society of America* **25**, 975-979 (1953)

[8] G.A. Miller and J.C.R. Licklider, "The intelligibility of interrupted speech", *Journal of the Acoustical Society of America*, **22**, 167-173, (1950)

[9] G.A. Miller and G.A. Heise, "The trill threshold", *Journal of the Acoustical Society of America*, **22**, 637-638 (1950)

[10] D. Deutsch, "Grouping mechanisms in music", in Deutsch, D. ed. *The Psychology of Music*, 3rd edition, Academic Press, Amsterdam, pp. 183-248 (2013)

[11] L.P.A.S. van Noorden, "Temporal coherence in the perception of tone sequences", Unpublished doctoral thesis, Technical University, Eindhoven, 1975

[12] R.M. Warren, *Auditory Perception: An Analysis and Synthesis*, 3rd edition, Cambridge University Press, Cambridge, 2008

[13] C.I. Petkov, K.N. O'Connor, and M.L. Sutter, "Encoding of illusory continuity in primary auditory cortex", *Neuron*, **54**, 153-165 (2007)

[14] L. Riecke, F. Esposito, B. Bonte, and E. Formisano, "Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex", *Neuron*, **64**, 550-561 (2009)

[15] Y. Nakajima, T. Sasaki, K. Kanafuka, A. Miyamoto, G. Remijn, and G. ten Hoopen, "Illusory recouplings of onsets and terminations of glide tone components", *Perception & Psychophysics*, **62**, 1413-1425 (2000)

[16] T. Sasaki, Y. Nakajima, G. ten Hoopen, E. van Buuringen, B. Massier, T. Kojo, T. Kuroda, and K. Ueda, "Time-stretching: Illusory lengthening of filled auditory durations", *Attention, Perception, & Psychophysics*, **72**, 1404-1421 (2010)

[17] G.B. Remijn and Y. Nakajima, "The perceptual integration of auditory stimulus edges: An illusory short tone in stimulus patterns consisting of two partly overlapping glides", *Journal of Experimental Psychology: Human Perception and Performance*, **31**, 183-192 (2005)

[18] T. Kuroda, Y. Nakajima, S. Tsunashima, and T. Yasutake, "Effects of spectra and sound pressure levels on the occurrence of the gap transfer illusion", *Perception*, **38**, 411-428 (2009)

[19] T. Kuroda, Y. Nakajima, and S. Eguchi, "Illusory continuity without sufficient sound energy to fill a temporal gap: Examples of crossing glide tones", *Journal of Experimental Psychology: Human Perception and Performance*, **38**, 1254-1267 (2012)

Note: The audio files for the auditory demonstrations mentioned in this paper are available via the Acoustics Australia website for downloading this paper.