Human Auditory Mechanisms

Makio Kashino†

Abstract

What we hear does not simply correspond to the physical properties of the sounds reaching our ears. In many cases, auditory perception changes systematically depending on the preceding and subsequent sounds. Under certain conditions, even physically missing sounds can be perceived as if they were present. An understanding of the characteristics of auditory perception and the underlying brain mechanisms involved will provide clues for improving tele-presence and virtual reality technologies.

1. Introduction

What we perceive does not simply correspond to the properties of the physical stimuli reaching our sensory organs. In hearing, a sound may be perceived quite differently depending on the sound that precedes or follows it. Under certain conditions, even physically missing sounds can be perceived as if they were present. Processing acoustic information without taking the context dependency of perception into account may severely degrade the reality of the perceived sounds. On the other hand, by fully exploiting the characteristics of perception, it may be possible to achieve efficient information processing, including significant data reduction, without spoiling the perceived reality.

We are studying the dynamic context dependency of auditory perception. This paper introduces two auditory illusions that demonstrate dynamic context dependency: the perceptual restoration of physically missing sounds and the expansion and contraction of auditory space.

2. Perceptual restoration of physically missing sounds

2.1 Perceived only when masked

When portions of a recorded utterance are replaced by gaps of silence from 5 to 15 times a second, the utterance sounds disrupted and unnatural, and it is very difficult to understand what is being said. When the gaps are filled with broadband noise that is louder than the recorded voice, the utterance sounds more natural and continuous, making it much easier to understand (Fig. 1). In both cases, exactly the same amount of the speech signal has been deleted; however, the deleted portions are restored perceptually only when the gaps are filled with broadband noise [1].

This perceptual restoration effect clearly indicates that the sounds we hear are not copies of physical sounds. The brain fills in the sounds that should exist in the portions masked by noise bursts based on the information in the remaining speech signal. What we perceive is the result of such unconscious interpretation.

Our daily life in this noisy world would be quite inconvenient without the perceptual restoration function. For example, a slamming door may completely mask a portion of an utterance. Perceptual restoration compensates for the effect of this masking.

The perceptual restoration effect is not restricted to speech. It comes into play for various types of sounds, including music, environmental sounds, and pure tones, provided certain conditions are met. For the effect to occur, the acoustic (spectral, temporal, and spatial) characteristics of the interrupting sound must be sufficient to mask the interrupted sound if the two sounds were presented simultaneously (Fig. 2) [2], [3]. This is quite reasonable considering every-

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: kashino@avg.brl.ntt.co.jp
URL: http://www.brl.ntt.co.jp/IllusionForum/index2.html.

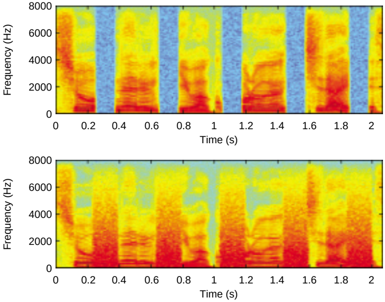

Fig. 1. Sound spectrograms of acoustic signals demonstrating perceptual restoration of missing portions of speech. Red indicates high energy; blue indicates low energy. Upper panel: Portions of Japanese utterance “Sho energii wa kokorogake shidai desu (The saving of energy depends on your motivation).” are regularly replaced by gaps of silence. Lower panel: Gaps in upper panel are filled with broadband noise. Both versions contain exactly the same amount of acoustic data for the original utterance, but intelligibility is drastically different.

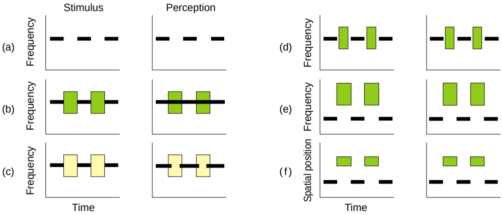

Fig. 2. Neccessary conditions for perceptual restoration. (a) A sequence of discrete tones is perceived as it is. (b) When the gaps in (a) are filled with noise bursts having an appropriate spectrum and sound pressure level, the tone is perceived as if it were continuous. (c) When the sound pressure level of the noise bursts is too low, the restoration is incomplete, even if the spectrum of the noise bursts is appropriate. (d) Noticeable gaps between the tone and noise burst prevent restoration. (e) When the frequency regions of tones and noise bursts are different, restoration does not occur. (f) When the spatial positions of the tones and noise bursts are different, restoration is suppressed even if other conditions are similar to those of (b).

day situations. Perceptual restoration of a segment is appropriate when that segment is actually present but masked by an extraneous sound. If the segment could not have been masked by the extraneous sound, then synthesizing the segment would be inappropriate. The selectivity of perceptual restoration reduces the possibility of an inappropriate perception of a signal fragment.

2.2 Reversal of past and future

The perceptual restoration effect is intriguing from the viewpoint of time. When a 200-ms gap of silence is introduced into an upward frequency glide, the gap is clearly noticeable (Fig. 3(a)), but when the gap is

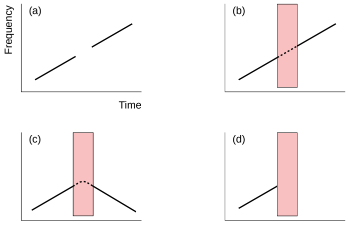

Fig. 3. Reversal of past and future in perceptual restoration effect.

filled with an appropriate noise burst, the perceptual restoration effect occurs and the glide is perceived as continuous again (Fig. 3(b)). When the noise burst is followed by a downward glide, listeners perceive a continuous glide changing its frequency smoothly (Fig. 3(c)). When no glide follows the noise burst, the glide is perceived as ending at the onset of the noise burst (Fig. 3(d)).

These phenomena are strange because no information is provided to listeners concerning what to restore in the portion masked by the noise burst until the glide following the noise burst appears. Nevertheless, listeners perceive a continuous glide with a smoothly changing frequency even when the noise is present. The same is true in speech. In an experiment in which a segment of a continuous utterance was replaced with a noise burst, we found that the segment was restored according to not only the speech features preceding the noise burst but also those following it [4]. The only plausible interpretation for these phenomena is that perceptual restoration occurs retroactively, after the information following the missing segment is provided. Physical time goes from the past to the future, but perceptual time apparently does not always correspond to the physical one.

2.3 Paradox of temporal acuity

Another intriguing aspect of perceptual restoration from the viewpoint of time is that, when a phoneme is deleted in an utterance and replaced with an appropriate noise burst, it is very difficult for the listener to locate the position of the noise burst. Indeed, most listeners mislocalize its position by several phonemes. It is very easy, of course, to judge the order of phonemes. In common-sense terms, a stopwatch that can measure time with an accuracy of 0.1 s, for example, can measure 1 s without any problem. In human auditory perception, on the other hand, it can be difficult to judge the coarse temporal order of acoustic events under certain conditions but easy to judge a much finer one.

In short, temporal order judgment is accurate for a sequence of related items, but poor for a sequence of unrelated items. This is sufficient for daily life because, rarely does a listener have to judge the fine timing of an irrelevant noise burst imposed on an utterance.

All of this means that the perceived timing of events does not directly reflect their physical timing; instead, it is created by the brain in the span of a few hundred milliseconds. Simply because the perceived timing differs from the physical one, however, does not mean that our perception is inaccurate. The auditory system is capable of restoring missing information and judging the order of relevant events appropriately.

3. Expansion and contraction of auditory space

3.1 Mechanism of sound localization

Another illusion that demonstrates the dynamic context dependency of auditory perception is the expansion and contraction of auditory space due to the preceding sound. Before going into the illusion, it will be helpful to summarize the mechanism of sound localization [5].

There are several cues that people use to localize a sound source. One of the major ones for horizontal sound localization is the difference in the times at which sound waves arrive at the ears. This is called the interaural time difference. If a sound source is to the right of a listener, for example, the sound wave from the source arrives at the right ear first. The interaural time difference is less than a millisecond. Another major cue is the difference in the sound pressure level at the ears, which is called the interaural level difference. The interaural level difference is created because a sound wave is attenuated by the shadow of the head. These two cues are primarily used for azimuthal localization. The main cues for judging elevation or distance are the spectral changes produced by the listener's head and pinnae, which depend on the sound source location.

Neural mechanisms for azimuthal sound source localization have been extensively studied using barn owls and cats. The neural processing revealed so far can be outlined as follows. Sound waves reaching the ears are first decomposed into many frequency channels. Then the signals in each frequency channel from ears are compared in the brainstem and the interaural time differences and level differences are extracted. These two types of cues for azimuthal sound localization are processed separately by specialized neural mechanisms. At higher stages of neural processing, the information from the different frequency channels is integrated into a neural map representing sound location.

3.2 Systematic shifts in sound localization

Although much is known about the perceptual characteristics and underlying neural mechanisms for sound source localization, most studies have focused on a single source. In the real world, however, several sound sources often exist concurrently or successively at various positions. These sound sources may interact with each other in sound source localization. To clarify this point, we examined the effect of a preceding sound on the localization of a subsequent test sound [6]. We found that the perceived position of the test sound shifts away from the position of the preceding sound, as if the perceptual space had been expanded in the neighborhood of the preceding sound (Fig. 4). The shift was the largest (as large as 15 to 20 degrees) when the two sounds were separated by 40 degrees.

Further experiments revealed that this illusion occurred only when the two successive sounds were close in frequency. Moreover, the independent manipulation of interaural time and level differences was found to produce nearly independent effects. Therefore, this illusion presumably involves neural mechanisms that process interaural time and level differences separately for each frequency channel in the early stages of binaural processing.

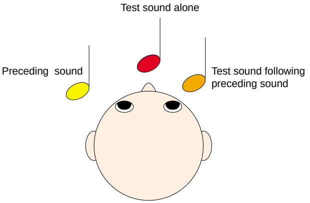

Fig. 4. Shift in perceived sound location due to preceding sound. A test sound that is perceived at the center without a preceding sound may be perceived as if it shifted to the right immediately following the presentation of a sound at the left.

3.3 Computational model of the illusion

Is it possible then to explain the perceptual characteristics of the illusion based on the mechanisms by which interaural time and level differences are extracted? Here, we focus on the interaural time difference. A hypothesis proposed more than 50 years ago to explain the neural mechanisms for detecting fine interaural time differences of less than 1 ms is still widely accepted today. The hypothesis, called the delay and coincidence circuit, is as follows. The signals from the ears are fed into a "coincidence detector" after a certain amount of transmission delay is applied to the signal from one side. If the coincidence detector is activated, it means that the interaural time difference between the sound waves is exactly canceled out by the internal delay. Thus, a set of coincidence detectors, each having a different amount of internal delay, can code interaural time differences according to which coincidence detector is activated most. Such a neural circuit was shown to exist in a brainstem nucleus of the barn owl in 1990. Today, a similar neural circuit is assumed to exist in other birds and mammals as well.

This circuit, however, cannot fully explain the illusion. Our idea is that the interaural time difference is coded not according to which coincidence detector is activated most, but by the pattern of excitation (for example, a centroid of many coincidence detectors. An essential feature of our model is that each coincidence detector changes its sensitivity according to the

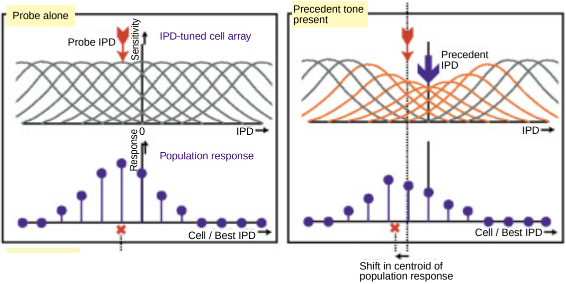

Fig. 5. Computational model of context dependency of auditory spatial processing. Left panel: Test sound alone. There is a population of neurons each tuned to a specific interaural time difference (indicated here as IPD) in the brain. The response pattern of the population of neurons to an interaural time difference (red arrow) is shown at the bottom. The centroid of the response pattern (red cross) corresponds to the perceived location. Right panel: An interaural time difference (red arrow) (the same as in the left panel) is presented immediately following a sound having an interaural time difference indicated by the blue arrow. Because the sensitivity of each neuron has been reduced due to the preceding sound, the centroid of the response pattern now shifts away from the preceding interaural time difference.

recent input in such a way that sensitivity is reduced for a large input and increased for a small one (Fig. 5). The computer simulation of this model fits the actual characteristics of the illusion well.

3.4 Neural basis of the illusion

To explore the neural basis of the computational model of the illusion, we conducted a series of animal experiments in which neural activity was recorded in the inferior colliculus, a brainstem nucleus that plays an important role in sound localization [7].

The experimental stimuli were sequences of two tones having various interaural time differences, like those used in our psychophysical experiments. Most neurons we measured were found to reduce their activity depending on the activity to the preceding sound. This is consistent with our computational model of the illusion.

3.5 Adaptive coding of acoustic information

Now, let us look at the functional significance of the context dependency of auditory spatial processing in the real world. For absolute judgment of sound localization, it would be disadvantageous to change the characteristics of acoustic cue coding adaptively depending on the input signals. Adaptive coding has, however, an advantage in information processing. The discrimination of the spatial positions of two sounds is improved in the neighborhood of a preceding sound, at the cost of degradation elsewhere [8]. The expansion of the perceptual space around the preceding sound is exactly like looking at things through a magnifying glass.

Generally speaking, the firing rate of each neuron represents an input value within a certain range. The range is not very wide; the firing induced by the input saturates at the top end and is masked by intrinsic noise at the bottom end. On the other hand, a neuron may be able to code a wider range of input effectively by changing its sensitivity adaptively according to the input value, as assumed in our model. If each neuron behaves like this, the difference in the activity patterns of the entire population of neurons becomes larger for slightly different input values near the recently experienced input value, resulting in better discrimination.

Our findings show that the information processing of auditory space is more dynamic and adaptive than has previously been assumed. To process information more efficiently, the brain continuously changes the focus of processing based on the input signal.

4. Conclusion

The two types of illusions described in this article clearly indicate that the sounds we perceive are not copies of physical sounds. The brain continuously produces hypotheses about acoustic events and looks for evidence for the hypotheses in the sounds coming into the ears. The world we perceive is nothing but the most plausible interpretation of the input signals. In the course of auditory information processing, the properties of neural systems change, and the important information is selected. Such dynamic context dependency enables a listener to grasp what is occurring where more efficiently and stably in a world in which various events take place successively.

An understanding of the characteristics and mechanisms of auditory information processing in the brain will be useful in developing technologies to recode and reproduce vast amounts of ever-changing acoustic information efficiently and with a sense of reality. This will lead to improved tele-presence and virtual reality systems.

References

[1] R. M. Warren: "Perceptual restoration of missing speech sounds," Science, Vol. 167, pp. 392-393, 1970.

[2] R. M. Warren: "Perceptual restoration of obliterated sounds," Psychological Bulletin, Vol. 96, pp. 371-383, 1984.

[3] M. Kashino and R. M. Warren: "Binaural release from temporal induction," Perception & Psychophysics, Vol. 58, pp. 899-905, 1996.

[4] M. Kashino: "The perception of Japanese intervocalic stop consonants based on distributed cues preceding and following stop closure," The Journal of the Acoustical Society of Japan, Vol. 48, pp. 76-86, 1992.

[5] Y. E. Cohen, and E. I. Knudsen: "Maps versus clusters: different representations of auditory space in the midbrain and forebrain," Trends in Neuroscience, Vol. 22, pp. 128-135, 1999.

[6] M. Kashino and S. Nishida: "Adaptation in sound localization revealed by auditory aftereffects," Journal of the Acoustical Society of America, Vol. 103, pp. 3597-3604, 1998.

[7] S. Furukawa, K. Maki, H. Riquimaroux, and M. Kashino: "Sensitivity of Inferior-Collicular Neurons to Interaural-Phase Difference (IPD) in the Presence of Preceding Sound with Various IPD," ARO Mid-Winter Meeting, Vol. 47, 2002.

[8] M. Kashino: "Adaptation in sound localization revealed by auditory after-effects," In A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Eds.), Psychological and physiological advances in hearing. London: Whurr Publishers, pp. 322-328, 1998.

Makio Kashino

Senior Research Scientist, Supervisor, Group Leader, Sensory and Motor Research Group, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received B.A., M.A. and Ph.D. degrees in experimental psychology from the University of Tokyo. In 1989, he joined NTT Basic Research Laboratories, Tokyo, Japan. His expertise is in psychophysics and cognitive neuroscience, especially of auditory perception. Dr. Kashino is a member of the Acoustical Society of America, the Acoustical Society of Japan, the Association for Research in Otolaryngology, and the Japanese Psychological Association.