INTERACTIONS AND INTEGRATIONS OF MULTIPLE SENSORY CHANNELS IN HUMAN BRAIN

Shin'ya Nishida

NTT Communication Science Laboratories, NTT Corporation, Atsugi, Japan

ABSTRACT

Here I describe a couple of new principles with regard to interactions and integrations of multiple sensory channels in the human brain. First, as opposed to the general belief that the perception of shape and that of color are relatively independent of motion processing, human visual system integrates shape and color signals along perceived motion trajectory in order to improve visibility of shape and color of moving objects. Second, when the human sensory system binds the outputs of different sensory channels, (including audio-visual signals) based on their temporal synchrony, it uses only sparse salient features rather than using the time courses of full sensory signals. We believe these principles are potentially useful for development of effective audiovisual processing and presentation devices.

1. INTRODUCTION

Human sensory system consists of multiple parallel channels for different sensory modalities (e.g., vision and audition). Each modality is further subdivided into processing channels for different attributes (e.g., color, shape and motion in vision). If the whole sensory system is merely an aggregation of independent channels for different sensory attributes, for the purpose of understanding the system and finding effective human perceptual factors for audiovisual processing and presentation technologies, one can simply focus on specific sensory attributes one by one. However, the truth is that various sorts of cross-channel interactions play essential roles in human sensory processing.

Here I describe a couple of new principles of cross-channel interaction that have been revealed by our recent studies. One is that the human visual system uses motion signals to improve visibility of shape and color of moving objects. The other is that the human sensory system uses only sparse salient features to find temporal synchrony between the outputs of different channels. We believe these principles are not only neuroscientifically important, but also potentially useful for development of effective audiovisual coding and presentation devices.

2. MOTION-BASED SHAPE AND COLOR PERCEPTION

2.1. Motion-based shape perception in multi-slit view

It is widely believed that the human visual cortex initially analyses spatial shapes independent of their movements. A phenomenon that challenges this notion of separate processing is spatiotemporal pattern interpolation [1, 2]. Consider an array of vertical slits, each having one pixel width. The slit separation is so wide that a stationary pattern shown through the slit array is hard to recognize. However, when a pattern is horizontally moving behind the stationary slit array, it can be effortlessly recognized in spite of only a fraction of the image being actually presented at any instant of time.

In a series of psychophysical experiments, I [3] examined whether the spatiotemporal pattern interpolation effect is supported by motion mechanisms. The results indicate that multi-slit viewing cannot be ascribed to motion-irrelevant factors such as retinal painting by tracking eye movements. Pattern perception is more strongly impaired by the masking noise moving in the same direction than by the noise moving in the opposite direction, which indicates direction-selectivity of the pattern interpolation mechanism. A direction-selective impairment of pattern perception by motion adaptation also indicates direction-selectivity of the interpolation mechanism. Finally, the map of effective spatial frequencies, estimated by a reverse-correlation technique, indicates observers' perception of higher spatial frequencies, the recovery of which is theoretically impossible without the aid of motion information. These findings provide clear evidence against the notion of separate analysis of pattern and motion. The visual system uses motion mechanisms for integration of spatial pattern information along the trajectory of pattern movement. This is a useful mechanism for the visual system to obtain clear shape perception of moving patterns (**Fig. 1**).

2.2. Motion-based color mixing and segregation

As in the case of shape perception, although it is generally thought that color perception is relatively independent of motion processing, the human visual system also uses

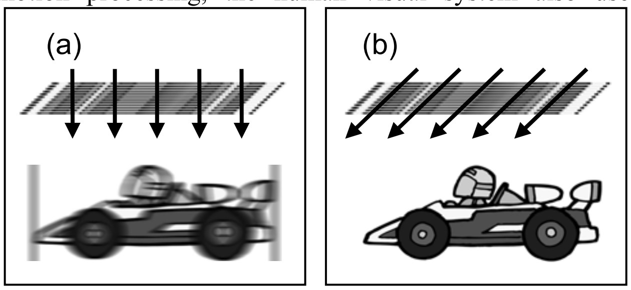

Fig. 1. Temporal integration of shape and color signals can improve the signal-to-noise ratio of the visual processing. However, when signals coming from moving objects are simply integrated at each position, motion blur will degrade the image quality. Our findings suggest that the human visual system obtains clear images of the moving objects by means of integrating shape and color signals along the trajectories of motion.

motion information to achieve veridical color perception for moving objects. Specifically, we found a new illusion in which moving bars that alternate between two colors (e.g., red and green) are perceived as the mixed color (yellow) even though the separate colors always stimulate different parts of the retina [4]. This motion-induced color mixing cannot be ascribed to the spatiotemporal resolution limit of the visual system or by involuntary eye movements of the observer. This finding indicates the existence of a neural mechanism that integrates color signals along a motion trajectory. This mechanism may contribute to the veridical perception of moving colors without motion blur (**Fig. 1**).

In agreement with this hypothesis, we also found motion-induced color segregation in which temporal alternations of two colors on the retina are perceptually segregated more veridically when they are presented as moving patterns rather than as stationary alternations at the same rate [5].

2.3. Integration of shape and color along motion trajectory: Implications for application

Temporal integration of visual signals is a critical mechanism of the human visual system that engineers have to consider for specification of the methods to code and present motion pictures. Previous researchers have only considered passive temporal signal integration at the same retinal position. They believe that modulations of a temporal frequency higher than the temporal resolution is just invisible, but the truth is not so simple. Our finding of active signal integration along trajectory of the object (with and without eye movements) could lead to effective presentations of motion pictures, as demonstrated by the multi-slit display. We expect the obtained method will be free from the problem arising from the assumption of passive temporal integration, including color breaking. In addition, the fact that the human visual system is relatively insensitive to temporal variations within a trajectory of motion may be applicable to movie image compression.

Note also that one of our findings (motion-induced color segregation) suggests that the standard method of using flicker fusion may underestimate the upper temporal resolution of color perception. The upper resolution of color perception measured using moving stimuli could be higher than might otherwise be expected.

The cross-channel interactions described in this section are supposed to take place at relatively early sensory levels. Consider next is how the outputs of sensory channels are combined at later perceptual levels.

3. SALIENCY-BASED CROSS-CHANNEL TEMPORAL INTEGRATION

3.1. Apparent asynchrony of color and motion

The transmission and processing of sensory information by neural mechanisms takes time, and the amount of time taken varies significantly across channels. As a result, sensory signals referring to different aspects of a single object or event are spatially and temporally spread over the brain's neural network. To perceive a coherent world, the brain has to group these signals in a way that allows them to reference a single object or event. An important physical constraint on the grouping of signals is the temporal coincidence of the physical attributes to which they refer.

A phenomenon that suggested us the principle underlying cross-channel temporal comparison is color-motion asynchrony. This is an illusion in which temporal judgments about the relationship between rapid alternations of color and motion direction indicate a large apparent delay of motion relative to color [6]. This illusion was originally interpreted as indicating that the motion channel requires longer processing time than does the color channel [6, 7]. This processing time hypothesis however is theoretically problematic in that it has the logical shortcoming of identifying subjective temporal relationships with the time course of objective neural events. In addition, this account assumes a "brain time" mechanism that is poorly designed in the sense that processing delay is added to "event time" estimation.

The alternative hypothesis that we proposed is that color-motion asynchrony results from matching inappropriate time markers. That is, color change is matched with more salient position change (motion) rather than with less salient motion direction change. We consider that cross-channel temporal comparison is based on the comparison of time markers by a mid-level perceptual process. The time markers for cross-channel comparison are amodal tokens, each linked to a specific temporal event.

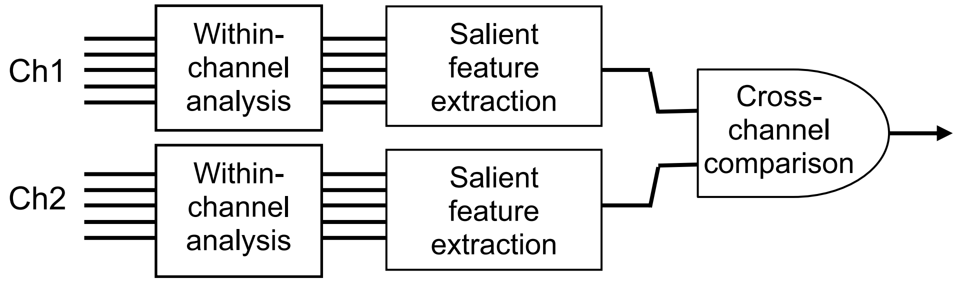

Fig. 2. Basic structure of our model of cross-channel temporal comparison based on salient feature matching.

This representation is supposed to be based on salient, figural, features extracted from early-level sensory signals. A temporal marker should reference the time a specific event occurs in the world rather than the time the processing of the event completes in the brain.

There are several findings that are consistent with our time marker theory, but not with the processing time hypothesis. First, it is the temporal structure difference not the attribute difference that determines the occurrence of perceptual asynchrony [8]. Second, perceptual asynchrony depends on the temporal properties of the stimulus — it occurs only for moderately rapid alternation rate [8, 9]. Third, perceptual asynchrony is not accompanied by a corresponding difference in reaction time [8]. Fourth, color-motion asynchrony disappears when alternating motion generates a percept of motion transparency [10, 11].

3.2. Audio-visual synchrony perception

To test the applicability of our time marker theory (**Fig. 2)** to cross-channel temporal comparisons in general), we analyzed the mechanism of audio-visual synchrony perception.

If audio-visual synchrony is also detected by a slow mid-level process, the temporal resolution for the detection of audiovisual synchrony is expected to be low. Indeed, for a repetitive pulse train, the upper limit is only ~4Hz [12]. For the higher frequencies, while the subjects can clearly perceive visual changes and auditory changes, they cannot determine their temporal relationship.

To be exact, this temporal limit should be evaluated not in terms of temporal frequency of stimulus modulation, but in terms of density of stimulus features [13]. Even when the stimulus contains low-frequency energy sufficient to support discrimination, synchrony discrimination is not possible for dense random pulses. Furthermore, even when the stimulus itself is dense, synchrony discrimination is possible if it includes sparse salient features (such as red pulse in white pulse, high pitch pulse in low pitch pulse). The features selected by top-down attention are also useful for audiovisual synchrony discrimination. These results suggest that audio-visual synchrony perception is based on the matching of salient features extracted from each sensory modality.

That audio-visual synchrony detection is mediated by a mid-level attentive process, rather than by an early pre-attentive process, is further supported by the finding that visual search for a target that changes in synchrony with an auditory stimulus becomes difficult as the number of uncorrelated visual distractors increases — the typical pattern one finds in serial search [14].

These properties of audiovisual synchrony perception are in agreement with those of spatial, spatiotemporal and temporal binding between different attributes within vision [15-17], which supports our hypothesis that saliency-based time marker matching is a common principle in mid-level temporal binding.

3.3. Recalibration of audio-visual simultaneity

For veridical estimation of event time, we think it is likely that the brain extracts timing signals (time markers) from early evoked neural responses. However, there is a problem for this strategy in the case of audiovisual binding, since there are time differences in physical transmission and sensory processing between the two modalities.

One strategy our brain might take to overcome this difficulty is to adaptively recalibrate the simultaneity point from daily experience of audio-visual events. Our study [18] found novel psychophysical adaptation effects in which exposure to a fixed audio-visual lag for several minutes shifts the centre of subjective simultaneity responses in the direction that would reduce the adapted lag. This finding suggests that the brain attempts to adjust subjective simultaneity across different modalities by reducing constant lags between inputs that are likely to arise from the same physical events. It also indicates that the existence of early temporal discrepancies between channels is not a critical problem for the time marker theory.

3.4. Cross-channel temporal integration and recalibration: Implications for application

Considering dubbing of foreign movies and animations, one would not be surprised at our proposal that apparent audio-visual synchrony does not require exact matching of signal details. However, there was no formal understanding of the perceptual mechanism underlying audio-visual synchrony perception, nor the recognition of the similarity with the other cross-channel synchrony perception. Discovery of the principle of saliency-based cross-channel temporal integration could lead to the development of the system that automatically judges the quality of apparent synchrony for human users, given that one can successfully model how the human sensory system extracts salient temporal features from the stream of sensory signals.

The recalibration of audio-visual simultaneity discredits general belief that the range of tolerance of audio-visual time lag is predetermined. It could vary among individuals or even within individuals depending on the surrounding audio-visual environment. This finding provides substantial guidelines in technological development of man-machine interfaces, multimedia, virtual reality, internet education tools, etc. where auditory and visual signals may or may not be emitted from stimulators in perfect synchrony. More specifically, this finding might be applicable to reduction of audio-visual asynchronies such as those taking place in teleconference systems.

4. CONCLUSION

The human sensory system involves many channels. As reviewed in this paper, investigating interaction and integration across those channels is a fruitful research strategy to gain useful insight into development of effective audiovisual technologies.

5. REFERENCES

[1] J. Ross, "A new type of display relying on vision's sensitivity to motion [proceedings]," *J Physiol*, vol. 271, pp. 2P-3P, 1977.

[2] D. C. Burr, "Acuity for apparent vernier offset," *Vision Research*, vol. 19, pp. 835-837, 1979.

[3] S. Nishida, "Motion-based analysis of spatial patterns by the human visual system," *Curr Biol*, vol. 14, pp. 830-9, 2004.

[4] S. Nishida, J. Watanabe, S. Tachi, and I. Kuriki, "Motion-induced colour mixture," presented at Vision Siences Society, Sarasota, 2004.

[5] S. Nishida, J. Watanabe, and I. Kuriki, "Motion-induced colour segregation," presented at Vision Siences Society, Sarasota, 2005.

[6] K. Moutoussis and S. Zeki, "A direct demonstration of perceptual asynchrony in vision," *Proc R Soc Lond B Biol Sci*, vol. 264, pp. 393-9, 1997.

[7] D. H. Arnold, C. W. Clifford, and P. Wenderoth, "Asynchronous processing in vision: color leads motion," *Curr Biol*, vol. 11, pp. 596-600, 2001.

[8] S. Nishida and A. Johnston, "Marker correspondence, not processing latency, determines temporal binding of visual attributes," *Curr Biol*, vol. 12, pp. 359-68, 2002.

[9] H. E. Bedell, S. T. Chung, H. Ogmen, and S. S. Patel, "Color and motion: which is the tortoise and which is the hare?," *Vision Res*, vol. 43, pp. 2403-12, 2003.

[10] C. W. Clifford, B. Spehar, and J. Pearson, "Motion transparency promotes synchronous perceptual binding," *Vision Res*, vol. 44, pp. 3073-80, 2004.

[11] F. Moradi and S. Shimojo, "Perceptual-binding and persistent surface segregation," *Vision Res*, vol. 44, pp. 2885-99, 2004.

[12] W. Fujisaki and S. Nishida, "Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals," *Exp Brain Res*, vol. 166, pp. 455-464, 2005.

[13] W. Fujisaki and S. Nishida, "Feature-based, post-attentive processing for temporal synchrony perception of audiovisual signals revealed by random pulse trains pulse trains," presented at 6th Annual meeting of the International Multisensory Research Forum, Robereto (Italy), 2005.

[14] W. Fujisaki, A. Koene, D. Arnold, A. Johnston, and S. Nishida, "Visual search for a target changing in synchrony with an auditory signal," *Proc R Soc Lond B Biol Sci*, vol. in press, 2005.

[15] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognit Psychol*, vol. 12, pp. 97-136, 1980.

[16] A. O. Holcombe and P. Cavanagh, "Early binding of feature pairs for visual perception," *Nat Neurosci*, vol. 4, pp. 127-8., 2001.

[17] H. Ashida, A. E. Seiffert, and N. Osaka, "Inefficient visual search for second-order motion," *J Opt Soc Am A Opt Image Sci Vis*, vol. 18, pp. 2255-66., 2001.

[18] W. Fujisaki, S. Shimojo, M. Kashino, and S. Nishida, "Recalibration of audiovisual simultaneity," *Nat Neurosci*, vol. 7, pp. 773-8, 2004.