Received: 13 June 2022 | Accepted: 17 August 2022

DOI: 10.1002/pchj.603

ORIGINAL ARTICLE

PsyCh Journal

Holistic temporal order judgment of tones requires top-down disentanglement

Zhilin Zhang[1,2,3,4] | **Xiaoxiong Lin**[1,5,6] | **Yan Bao**[1,4,7]

[1]School of Psychological and Cognitive Sciences, Peking University, Beijing

[2]Yuanpei College, Peking University, Beijing, China

[3]Department of Psychology, New York University, New York, USA

[4]Institute of Medical Psychology, Ludwig Maximilian University, Munich, Germany

[5]Graduate School of Systemic Neurosciences, Ludwig Maximilian University, Munich, Germany

[6]Department of Neurosurgery, Klinikum rechts der Isar der Technischen Universität München

[7]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing

Correspondence

Yan Bao, School of Psychological and Cognitive Sciences, Peking University, 5 Yiheyuan Road, Beijing, 100080, China.
Email: baoyan@pku.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 31371018, 31771213

Abstract

How temporal sequence gets organized is a central topic in cognitive processing. In a high-frequency time window of tens of milliseconds, the temporal order is reconstructed rather than mirroring the sequence of events objectively in physical time. Two separate phases or strategies, a holistic coding phase that groups successively presented events as a gestalt and a disentanglement phase that decodes the temporal order of discrete events from the gestalt representation, may presumably be involved in the perception of temporal order across different modalities. With a temporal order adaptation protocol of pure tones using glide adaptors, the present study demonstrated a dissociation between constant discriminability and shifted subjective simultaneity across different adaptor directions. While discriminability of temporal order was not adapted by glides, revealing a constant coding sensitivity of different asynchronies, the shift of subjective simultaneity indicated the recalibration of a top-down disentanglement of the holistic processing under the influence of glide adaptors. The results suggest a dual-phase holistic processing in temporal order perception, supporting two separate cognitive strategies for event timing on the sub-second level.

KEYWORDS

auditory perception, gestalt perception, temporal order perception

INTRODUCTION

Organization of temporal information in perception plays an important role as a logistic function for cognitively representing our dynamic world (Pöppel & Bao, 2014). As an elementary function in the hierarchical scales of temporal perception, the ability to accurately discriminate the order of two events in tens of milliseconds has been regarded as one quantum unit of the perception of time (Bao et al., 2015; Buhusi & Meck, 2005; Hirsh & Sherrick, 1961; Pöppel, 1970, 1997, 2009). There has long been debates on how temporal experience emerges from such a small time window, but the mechanism of how temporal order is perceived has still to be agreed upon (Fekete et al., 2018; Fingelkurts & Fingelkurts, 2006; Herzog et al., 2020; Holcombe, 2015; Lee, 2014; VanRullen & Koch, 2003; White, 2018). Several hypotheses, such as attention-based successive sampling and discrete temporal windows, have been raised to count on temporal properties of perceptual processes, which also underlie perception of temporal order itself (White, 2018).

The complementarity between objective, physical time and subjective, psychological time (e.g., Holcombe, 2015; Koenderink et al., 2012; Lee, 2014; Pöppel & Bao, 2014; Zhou et al., 2014) can be used as a conceptual frame for a general taxonomy to better understand the representation of temporal order in neural systems (Bao et al., 2017). Instead of mirroring what is happening physically in time, the neurocognitive machinery reconstructs sequential information of event time in a brief temporal interval. Mismatch between objective time and cognitive reconstruction results in failure of temporal order judgment. Two different mechanisms can presumably be derived from this reconstruction framework; the canonical mechanism relies on an analysis of discrete event registration, but an alternative mechanism can process temporally close events as a gestalt. With the first hypothesized mechanism, the perception of temporal order is determined by the neural registration of different events in time, which, however, may be shadowed by noise. Given the representational noise, two events need to be separated by a minimum temporal interval, known as temporal order threshold (TOT), for them to be registered correctly on internal time (Ulrich, 1987). This canonical view of sequential event detection is consistent with most observations, where the threshold for accurate perception of temporal order lies around 20 to 40 milliseconds of stimulus onset asynchrony (SOA; Babkoff & Fostick, 2013; Fostick & Babkoff, 2013; Kanabus et al., 2002; Liang et al., 2015; Ulrich, 1987), and this discrete time window has been regarded as the atomic unit of temporal processing (Pöppel, 1997). However, the registration theory for temporal order judgment is restricted to this relatively stationary discrete time window, and this mechanism is not sufficient to explain the greatly varied temporal order threshold spectrum in different modalities and under various experimental paradigms. Temporal order thresholds for pure tones, for example, can be much smaller than the TOT measured with other protocols (Bao et al., 2013, 2014; Fink et al., 2006; Fostick & Babkoff, 2013). A TOT can be around 10 ms, which is below the minimum values observed in several studies. Thus, another neural mechanism has to exist, other than the event detection and registration mechanism, that reconstructs temporal order from shorter intervals. It is proposed that a holistic perception may be responsible for this processing (Bao et al., 2013, 2014). A grouped temporal perception may provide temporal information in gestalt form to analyze temporal order. Time order perception may therefore go beyond the sensitivity limit by integrating different events that happen successively and successfully decipher the integration. Two events may not only be registered as two separate markers in the temporal processing stream, but they can also be represented as a holistic gestalt.

The holistic processing of temporal order is supported by evidence from various uni- or multi-modality results of asynchrony adaptation (Bennett & Westheimer, 1985; Cai et al., 2012; Okada & Kashino, 2003; Roach et al., 2011). With a certain order of asynchrony adaptors presented prior to the temporal order tasks, shift of point of subjective simultaneity (PSS) indicates a recalibration of subjective temporal order (Di Luca et al., 2009; Fujisaki et al., 2004; Miyazaki et al., 2006), revealing that two events separated by a small time interval can be processed as a holistic gestalt in temporal perception (Koenderink et al., 2012). This asynchrony adaptation effect exists across various modalities in the form of apparent motion with visual stimuli of different position (Bennett & Westheimer, 1985) and glide adaptation with auditory tones of different frequency (Okada & Kashino, 2003), while similar effects have also been observed with tactile (Miyazaki et al., 2006), multi-sensory (Fujisaki et al., 2004), and sensorimotor temporal order judgment (Cunningham et al., 2001; Stetson et al., 2006). In addition to behavioral results, this holistic processing of time has also been supported by neural correlates of glide-processing neurons (Fuzessery et al., 2006; Tian & Rauschecker, 2004), as well as cortices that represent long-range apparent motion, auditory temporal gestalt, or motion-projection processing of tactile temporal order (Notter et al., 2019; Takahashi et al., 2013; Zhuo et al., 2003). The neural correlates of holistic processing further support that a gestalt can be formed during the processing of temporally close events, which may serve as a prerequisite function of the strategic processing of extracting the temporal order of the events. Intuitively, these gestalts can help get rid of some internal noise that is involved in the discrete registration of temporal order, by becoming heuristics of judging temporal order. For example, it has been shown that tonal language environment modulates temporal order judgment and yields better performance with temporal order judgment of close-frequency tone pairs (Bao et al., 2013). The second (rising) and fourth (falling) tones in Chinese, a typical tonal language, are natural holistic heuristics for native speakers to easily understand the words without explicitly reflecting on the temporal order of lower frequency and higher frequency components of the phonetics. In addition, non-tonal language exposure of native tonal language speakers improves their performance with distant-frequency tone pairs, which is believed to be due to obtaining new heuristic gestalts from non-tonal language learning that help holistic processing of distant-frequency tone pairs (Bao et al., 2014).

However, it has not been addressed previously whether our perceptual reconstruction of temporal order requires an additional process to decipher the sequence of events from the gestalt representation retrospectively in holistic processing that encodes two events as a whole. The asynchrony of two events that are perceived holistically needs to be disentangled backwards into a sequential temporal experience of two separate stimuli. Previous discussions mostly focus on the coding manner when holistic processing is involved, but holistic processing itself does not guarantee the readout of the precise temporal order. There is supposed to be another process to reconstruct the temporal order by disentangling discrete events from the holistic processing. With a series of previous studies that examines the temporal order discriminability of pure tones with different frequency distances, two components, holistic and analytical strategies, are proposed to be involved in temporal order perception (Bao et al., 2013, 2014). With the holistic and analytical strategies as two hypothesized cognitive heuristics, our current study is intended to reveal that the successively presented two events are perceptually analyzed to be disentangled from the holistic processing. Moreover, we argue that the extent of holistic processing and the capability of disentanglement, both as parts of temporal order reconstruction, conjointly influence the performance difference between close and distant frequency tone pairs.

The existence of disentanglement processing from gestalt perception should be established if the subjective simultaneity shift due to adaptation is a pure criterion shift without change of sensitivity for distinguishing temporal order. Meanwhile, since better disentanglement from holistic processing leads to better temporal order discriminability, the extent of criterion shift is supposed to be larger where temporal order discriminability is higher, indicating an overall better capability of temporal order reconstruction. To characterize the disentanglement process that happens during temporal order reconstruction from holistic perception, we applied an asynchrony adaptation protocol using glide adaptors with temporal order perception of pure tones (Okada & Kashino, 2003) and observed how subjective simultaneity criterion and temporal order discriminability differed due to adaptation. In addition, tone pairs of different frequency distances were presented to reveal how capability of temporal order reconstruction and performance of temporal order judgment were associated.

METHODS

Participants

Sixteen university students from Beijing participated in the experiment. All participants (aged 18–26 years, average 21.63 years; 8 males and 8 females) were native Chinese speakers, and had experience of speaking English. All participants were right-handed with normal hearing. All participants were given written informed consent prior to data collection. The study was approved by the local ethics committee of the School of Psychological and Cognitive Sciences, Peking University.

Stimuli

Acoustic stimuli used in the experiments were divided into targets and adaptors. All stimuli were adjusted according to equal-loudness curve (Suzuki & Takeshima, 2004), and were enveloped by 5-ms onset and offset volume cosine ramps. Target stimuli, which participants were supposed to judge the temporal order of, were 20 ms (10 ms of volume plateau) of sinusoidal tones in pairs. Each pair of target stimuli had one lower-frequency tone and one higher-frequency tone. The frequencies of target tone pairs were either 700 Hz and 891.9 Hz (1/3 octaves apart, *close frequency* condition), or 500 Hz and 2520 Hz (2 + 1/3 octaves apart, *distant frequency* condition). In different trials, glides chirping upward (*adaptor up* condition) or downward (*adaptor down* condition), or pure tones (*adaptor control* condition) are used as adaptors. Glides in *adaptor up* and *adaptor down* conditions were 25 ms (15 ms of volume plateau) of linear frequency chirps starting and ending at the same frequencies as target tone pairs. Adaptors in adaptor control condition were to use pure tone lasting for 25 ms (15 ms of volume plateau) at the geometric mean frequency of two target stimuli.

All stimuli were generated by Psychtoolbox Toolbox for MATLAB (Brainard, 1997; Kleiner et al., 2007), and were presented binaurally to participants via a headphone (Audio-technica ATH-M40x). Volume was adjusted by each participant so that they could hear all stimuli clearly and comfortably.

Procedure

In each trial, participants first heard six repeats of a 25-ms adaptor. The ISI between each adaptor was 75 ms. Two paired target stimuli were then presented with different SOAs. Participants were asked to judge whether the lower or the higher frequency tone appeared first by key pressing. The next trial began 1000 ms after response. The temporal structure of each trial is shown in Figure 1.

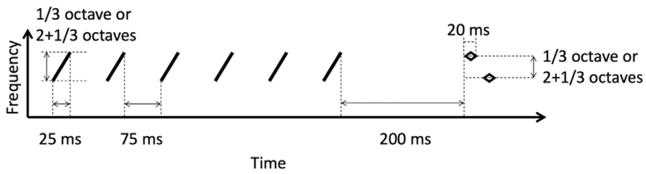

FIGURE 1 Temporal sequence within a single trial. Six repeats of an adaptor were presented prior to the target tone pair in each trial. In an actual trial, the adaptors could be upward (*adaptor up*), downward (*adaptor down*), or plain (*adaptor control*) according to the adaptor direction condition, and the frequency distances of tones could be 1/3 octave (*frequency close*) or 2 + 1/3 octaves (*frequency distant*). The six repeats of glides were always identical, and the starting and ending frequency were the same as the target tones in that trial.

All participants were tested for three adaptor direction conditions (*adaptor up, adaptor down*, and *adaptor control*), and two frequency distance conditions (*frequency close* and *frequency distant*). The adaptor direction conditions were intermingled trial by trial, but two frequency distance conditions were separated into two sessions carried out on two different days. Temporal order judgment was measured with different target stimuli SOAs using the method of constant stimuli. SOA <0 meant the higher frequency tone was presented first, and SOA >0 meant the lower frequency tone was presented first. When SOA = 0, two tones were presented simultaneously. In each condition, SOAs ranged from −60 ms to 60 ms in 5-ms step size. For each SOA, there were 20 repeated measurements. Each participant completed 3000 trials in two sessions.

The experiment was carried out in a quiet room. Participants first went through a practice block at the beginning of each session to get themselves acquainted with the task. They were given feedback of whether they answered correctly or incorrectly during the practice. There was no feedback in the formal experiment. Participants were told that their response time would not be recorded and that they needed to make the careful decision before they responded.

Data analysis

We used sigmoid Equation (1),

$$p(x) = \gamma + (1 - \gamma - \lambda)\frac{1}{1 + e^{-\beta(x-\alpha)}}, \tag{1}$$

to fit a psychometric function of participants' temporal order judgment. When x denoted SOA, $p(x)$ indicated the proportion of choosing "lower tone appears first". α and β were free parameters denoting position and slope of the curve, respectively, and γ and λ were lapse rates both set at fixed value 0.02. Psychometric curves were fitted using maximum likelihood estimation, and goodness of fit was evaluated by log-likelihood ratio of deviance scores (Wichmann & Hill, 2001). PSS was the SOA value where participants had equal probability of choosing "lower tone first" or "higher tone first." Just-noticeable difference (JND)

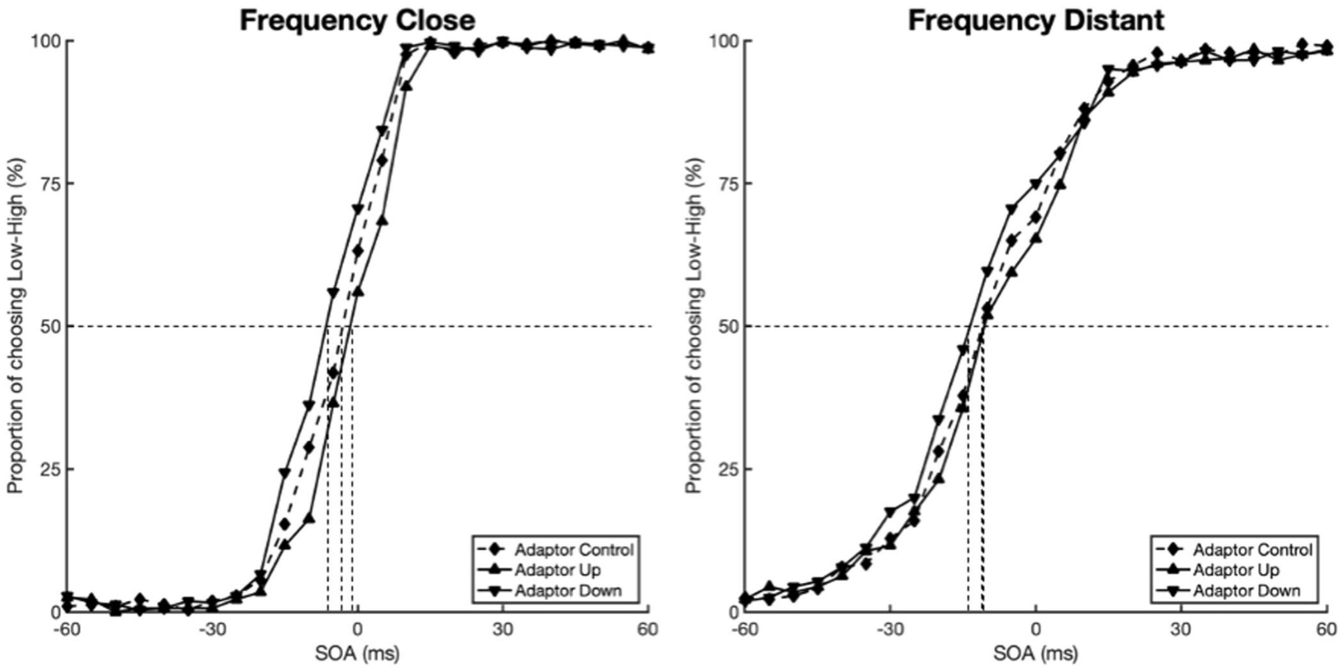

FIGURE 2 Group-averaged temporal order judgment performance. When the absolute stimulus onset asynchrony (SOA) value was small, task performance was close to chance level; when the absolute SOA was big enough, the proportion of participants who chose the right temporal order rose towards asymptotes, indicating that participants almost always made correct judgments. The slope of the rising phase indicates the temporal order discriminability: The steeper the rising phase is, the smaller SOA range in which uncertainty is shown, the better the ability of temporal order judgment is.

was half of the difference between SOA values that make $p(x)$ equal to 25% and 75%.

According to signal detection theory, under each frequency distance and adaptor direction condition, participants' temporal order judging criterion (c) and discriminability (d') were calculated by Equations (2) and (3):

$$c = -\frac{z(Hit) + z(False\ Alarm)}{2} \tag{2}$$

$$d' = z(Hit) - z(False\ Alarm). \tag{3}$$

Here, hit rate was defined by the proportion of choosing "lower tone appears first" when the actual presenting order was "low-high," and false alarm rate denotes the proportion of choosing "lower tone appears first" when actual temporal order of the tone pair was "high-low."

RESULTS

General performance pattern

The general task performance of psychometric curves is shown in Figure 2. There was an overall shift of psychometric function from *adaptor control* condition towards *adaptor up* or *adaptor down* condition, with *adaptor up* condition shifting to the right side and *adaptor down* condition shifting to the left. A clear shift between curves could be seen among the adaptor direction conditions in *frequency close* condition, while for *frequency distant* condition the three curves for different adaptor direction conditions were more entangled with each other. For the *frequency close* condition, proportion of choosing "low tone appears first" rose from zero to one quickly, indicating a smaller SOA range in which temporal order judgment is with higher uncertainty. This rising slope took much larger SOA range with *frequency distant* condition. The pattern agrees with the previous results that glide adaptors change subjective simultaneity (Okada & Kashino, 2003), and that the tested population is better at discriminating temporal order of tones with close frequency (Bao et al., 2013; Bao et al., 2014).

Simultaneity criterion

Simultaneity criterion shifted towards the positive side with the *adaptor up* condition and towards the negative with the *adaptor down* condition compared to the *adaptor control* condition, indicating a larger bias to choose "high-low" when adapted to upward glides and a larger bias to choose "low-high" when adapted to downward glides (Figure 3A). Criterion shift magnitude was larger with close frequency distance. There was a significant baseline bias under *adaptor control* condition of simultaneity criterion (*frequency close*: $t[15] = -2.540$, $p = .023$; *frequency distant*: $t[15] = -5.263$, $p < .001$). The bias was posited below zero, indicating that participants tended to judge the temporal order of the tone pairs as "low-high," even if the actual presenting order was simultaneous or "high-low." Shift from the *adaptor control* under the *frequency close* condition was significant

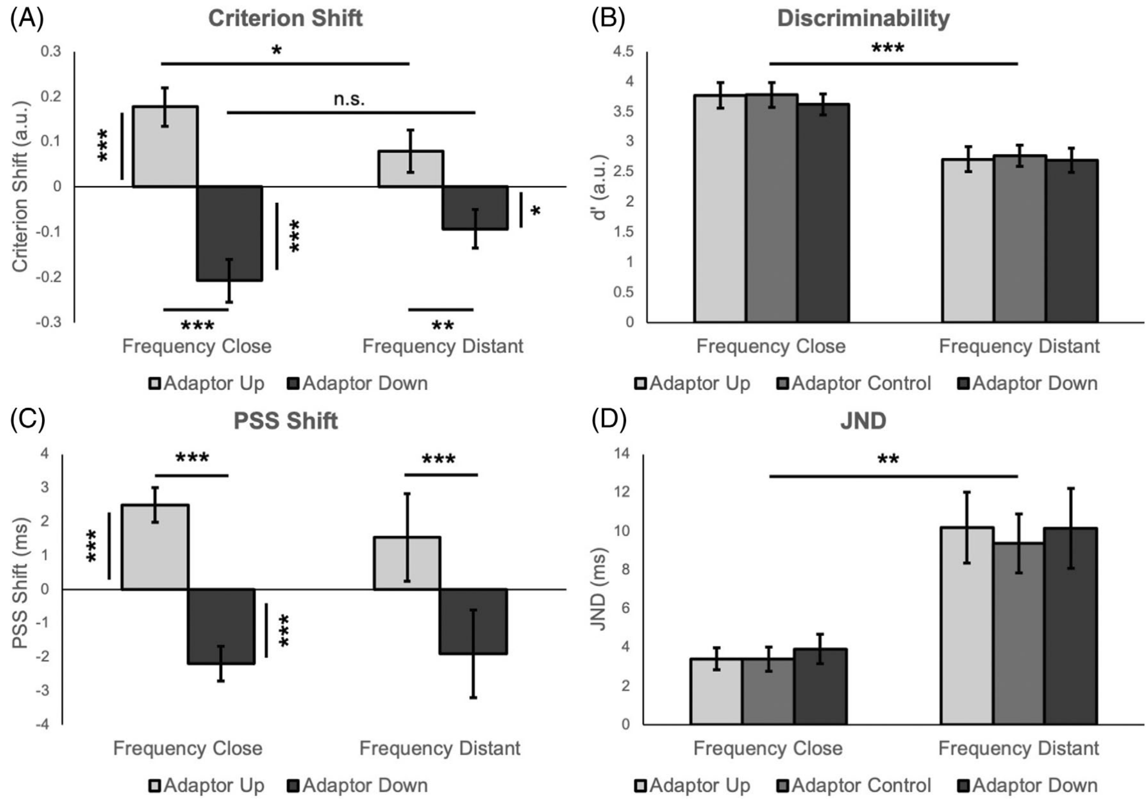

FIGURE 3 (A) Criterion shifted to opposite directions compared to *adaptor control* condition with adaptors of opposite directions. Magnitude of criterion shift was larger under *frequency close* condition than *frequency distant* condition. (B) Discriminability was not adapted and was higher under the *frequency close* than the *frequency distant* condition. (C) Point of subjective simultaneity (PSS) shift from the *adaptor control* condition was significant under the *frequency close* condition but not under the *frequency distant* condition. PSS shifted to the opposite directions with opposite adaptor directions. (D) Just-noticeable difference (JND) was not adapted, and *frequency close* condition has lower JND, indicating better temporal order judgment performance.

both under *adaptor up* (t[15] = 4.143, $p < .001$) and *adaptor down* (t[15] = −4.348, $p < .001$) conditions, while for the *frequency distant* condition, the shift was only significant under the *adaptor down* condition (t[15] = −2.182, $p = .045$) but not significant under the *adaptor up* condition (t[15] = 1.705, $p = 0.109$). A two-way (adaptor direction [up/down] × frequency distance [close/distant]) analysis of variance (ANOVA) of criterion shift revealed a significant interaction effect between frequency distance and glide direction (F[1, 15] = 7.514, $p = .015$, $\eta^2 = .775$). Criterion shift with upward glides was significantly higher than that with downward glides under both *frequency close* ($p < .001$) and *frequency distant* ($p = .002$) frequency conditions. Criterion shift under the *frequency close* condition was significantly lower than the *frequency distant* condition when glides were upward ($p = .038$), while downward glides led to no significant difference between frequency distance conditions ($p = 0.165$).

Discriminability (d′)

Discriminability was not adapted but was better under the *frequency close* condition than the *frequency distant* condition (Figure 3B). Unlike criterion, two-by-three (adaptor direction [control/up/down] × frequency distance [close/distant]) ANOVA showed that d' did not change significantly across the different adaptor directions (F[2,30] = 1.660, $p = 0.225$, $\eta^2 = .192$), but was significantly higher with the *frequency close* condition than the *frequency distant* condition (F[1,15] = 56.903, $p < .001$, $\eta^2 = .791$). There was no significant interaction between adaptor direction and frequency distance (F[2,30] = 0.422, $p = 0.664$, $\eta^2 = .057$).

Point of subjective simultaneity

PSS of the *adaptor up* or *adaptor down* condition shifted from the *adaptor control* condition due to adaptation, but the magnitude was larger under the *frequency close* condition than the *frequency distant* condition (Figure 3C). Similar to criterion, there was a significant baseline bias with PSS (*frequency close*: t[15] = −2.458, $p = .027$; *frequency distant*: t[15] = −4.936, $p < .001$). Moreover, the extent to which PSS baselines were biased was significantly larger with the *frequency distant* condition (F[1,15] = 12.042, $p = .003$, $\eta^2 = .445$). PSS shift from *adaptor control* to *adaptor up* or *adaptor down* condition was significantly deviant from zero under the *frequency close* condition (*adaptor up*: t[15] = 4.825, $p < .001$; *adaptor down*: t[15] = −4.241, $p < .001$) and not significantly different from zero under the *frequency distant* condition (*adaptor up*: t[15] = 1.196,

$p = .250$; *adaptor down*: $t[15] = -1.479$, $p = 0.160$). A two-way (adaptor direction [up/down] × frequency distance [close/distant]) ANOVA was applied to show the effects of frequency distance and adaptor direction. PSS shifted significantly to opposite directions between different adaptor direction conditions ($F[1,15] = 61.137$, $p < .001$, $\eta^2 = .803$), while the main effect of frequency ($F[1,15] = .081$, $p = 0.780$, $\eta^2 = .005$) and interaction effect ($F[1,15] = 0.835$, $p = 0.375$, $\eta^2 = .053$) were both non-significant.

Just-noticeable difference

Similar to discriminability, JND was not adapted but was smaller under the *frequency close* condition than the *frequency distant* condition, indicating an overall better temporal order judgment performance with close-frequency tones (Figure 3D). There was a significant main effect of frequency distance for JND ($F[1,15] = 18.006$, $p = .001$, $\eta^2 = .546$). No significant effect was found with main effect of adaptor direction ($F[2,30] = 1.476$, $p = 0.245$, $\eta^2 = .090$) or interaction effect ($F[2,30] = 0.266$, $p = 0.768$, $\eta^2 = .017$).

DISCUSSION

The present study reveals that the simultaneity criterion of temporal order judgment of auditory pure tones can be shifted by glide adaptors of different directions, while the discriminability is constant across different adaptor directions. This aftereffect caused by the glides exists in temporal order perception of tone pairs regardless of whether the frequency distance is close or far. However, the magnitudes of simultaneity shift can be deviant from each other, associated to different judgment performances. With close-frequency tone pairs, where temporal order discriminability is better, criterion shift is larger compared to distant-frequency tone pairs. The dissociation of discriminability and criterion shift under the influence of adaptor direction indicates that there are two separate processes of temporal order perception. We argue that temporal order perception is based on a dual-phase reconstruction process, with successive events registered holistically during the first phase, and then disentangled out of the holistic representation during the second phase. Temporal order discriminability and simultaneity shift by adaptation jointly indicate the capability of reconstructing temporal order.

In addition to providing converging evidence of a holistic processing of temporal order revealed by the asynchrony adaptation effect, the current study implies a second phase in temporal order perception, that is, to disentangle the integrated temporal synthesis back to representations of temporal order of discrete physical events. The disentanglement phase of deciphering holistic registration of temporal events does not contradict a potential neural mechanism of glide-sensitive neurons and their suspected adaptation. A hybrid mode of holistic encoding and analytical disentanglement is apparently operative in the perception of the temporal order of tones. While two events happen temporally close to each other, they are perceived holistically and then disentangled back into discrete events. In addition to the holistic perception, an extra step of event identification is crucial to the process. In the current protocol, different tones are coded in a glide-like manner, and these glide-like holistic representations can be decoded back into sequentially presented tone pairs, with "rising" to be decoded as "low-high" and "dropping" to be decoded as "high-low." This latter process of decoding is not addressed in previous studies, but it actually plays the critical role in reconstructing temporal organization in the psychological world. Our finding that subjective simultaneity criterion shifts while discriminability keeps the same across different adaptor directions distinguishes between the two phases of temporal order reconstruction. While sensibility to different directions of asynchronies depicts the ability to coding the stimuli holistically during the first phase, simultaneity criterion shift indicates the latter disentanglement phase. The fact that discriminability, indicated by d' or JND, does not get influenced by adaptors reveals an unchanged sensitivity for holistic coding, and the shift of subjective simultaneity, indicated by criterion or PSS, unfolded the adapted deciphering from the holistic representation. Meanwhile, the discriminability and simultaneity shift difference between close and distant frequency distances is due to the difference in capabilities of holistic processing and disentangling, respectively. For the native Chinese speakers in our study, the temporal order perception of close-frequency tone pairs is better than that of distant frequency tone pairs presumably due to language imprinting (Bao et al., 2013). The current results are consistent with this imbalanced advantage, showing better abilities for both holistic coding and disentangling of temporal order reconstruction.

During the reconstruction of temporal order, the disentanglement of event time takes place in a top-down manner where a deciphering process has a readout of temporal order from the holistic processing representation. However, when influenced by asynchrony adaptors, this shifted reconstruction should not be confused with a decision bias on a higher cognitive level. In spite of the top-down nature of the temporal order disentanglement, it still happens on an implicit and perceptual level, which is only delivered to higher level awareness after the disentanglement is completed. The participants in the study were instructed to report the phenomenological impression they had of the temporal order, and none of them noticed the distinction of adaptors among different adaptor direction conditions when questioned after they had finished all the sessions. Therefore, the possibility that the criterion shift is a response bias that happens during decision-making can be ruled out, since no decision was made based on condition changes that the participants were aware of. The idea that temporal order reconstruction is an automatic component in temporal order perception does not conflict with subjects' explicit experience of disentangling the sequence of stimuli out of a holistic strategy. It was typical among participants to report the close-frequency condition as "much easier" with a cognitive strategy. For close-frequency tones, many participants in this study were well aware that they could be grouped together to form a "rising" or "falling" trend and deliberately applied it as a strategy to temporal order judgment, while the similar explicit grouping strategy was much more difficult to apply with distant-frequency tone pairs, as was previously observed with native Chinese speakers (Bao et al., 2013). This introspective process actually implies that the concept of temporal order may be derivative from the holistic processing, and that sequence of events only emerges to consciousness when participants are instructed to report it explicitly in the experiment. In this sense, the disentangling phase of temporal order perception can even be regarded as artificial following the holistic processing of the originally separate events.

Although the proposed gestalt in our study can be regarded as a glide-like representation of the two separate tones, an alternative possibility that this gestalt is a group of two still-separate tones cannot be ruled out. Similar to visual ensembles, where a set of discrete stimuli can be grouped together to form a gestalt while they are not physically linked to each other, tones are not necessarily merged into a single "glide" to be holistically processed. Although it is intuitive to regard close-frequency tones as easier to be grouped into glides because of their smaller frequency distance, the same logic cannot stand when looking into temporal order judgment performance of tones with different frequency distances in different language groups. While native Chinese speakers are better at judging the temporal order of close-frequency tone pairs, native Polish speakers are better at distant-frequency tones (Bao et al., 2013). Although the tones are farther in distance, it is easier for the Polish speaking group to apply a holistic strategy. It has been proposed that Polish speakers are more exposed to the distant-frequency environment with the various consonants in their language, which leads to this performance pattern. In the current study, there is not enough evidence to show which kind of gestalt is involved, but it is still clear that a holistic processing is participating in the temporal order perception of pure tones.

The current reconstruction theory that temporal order information can be disentangled from holistic processing can be generalized to various modalities where asynchrony adaptation has been previously observed. Visual stroboscopic motion, auditory frequency gliding, and tactile reconstruction of spatial representation has been shown to underlie adaptation to visual, auditory and tactile temporal asynchrony (Bennett & Westheimer, 1985; Okada & Kashino, 2003; Takahashi et al., 2013). Meanwhile, compensation for a particular temporal asynchrony has also been observed for the temporal delay of multisensory and sensorimotor integrations (Cunningham et al., 2001; Fujisaki et al., 2004; Stetson et al., 2006). There is a universal temporal order reconstruction mechanism that is analogic across modalities, and such mechanism involves modulation to readouts of neural representations of time since the simultaneity shift is not accompanied by shift in time of stimulus-evoked neural signals (Holcombe, 2015; Roach et al., 2011). One possible framework that supports this universality of temporal order reconstruction may be the identity-related processing principle in temporal experience, where the decoding of gestalt perception of two events can be related to object identity and spatial representation (Takahashi et al., 2013; Zhou et al., 2014).

In this study, participants were able to tell the temporal order of two close-frequency tones even when they are largely overlapping with each other. The pattern deviates from the conventional threshold theory that an adequate interstimulus interval is needed for two stimuli to be distinguished from each other in a time sequence (Ulrich, 1987). This again supports that temporal information can be reconstructed out of internal representations as an alternative to registration according to physical time. Not only is this disentanglement crucial for perceptual time experience, but it may also be related to the distinction between simultaneity judgment and temporal order judgment (Miyazaki et al., 2016). While both are the most fundamental functions in the hierarchical perception of time, simultaneity judgment represents the ability to tell two events apart but not necessarily discriminate their order, and temporal order judgment requires a further step to correctly report their sequence (Pöppel, 1997). The point has been raised that the simultaneity judgment threshold, often regarded as the temporal resolution of conscious processing, can be independent with the temporal structure of consciousness, which involves more sequential representation (Herzog et al., 2020). The disentanglement process can be involved in the further organizing of perceptual sequence, which goes beyond the event registration that relies on the hypothesized resolution of temporal experience.

It is worth noticing that the reconstruction of temporal order out of holistic representation may apply specifically to high-frequency time windows lasting tens of milliseconds. While gestalt principle creates continuity among minimal time perceptual quanta, temporal perception on a larger time scale, such as 2 to 3 s, is still, to some extent, divided into relatively discrete "moments" for subjective experience of present to take place (Pöppel, 1997). Meanwhile, event detection and registration in discrete high-frequency time windows is still working in temporal order judgment of many kinds of stimuli, such as temporal order discrimination of monaurally presented clicks (Bao et al., 2013). The complementarity of continuity and discreteness in different mechanisms for time representation enriches the taxonomy in the hierarchy of time perception, indicating the intertwining involvement of discrete and continuous representation within and across different levels of time processing and time experience.

By addressing the disentanglement from holistic processing in temporal order perception, we emphasize the reconstruction of temporal information by special cognitive processes. The complementarities between bottom-up holistic registration of the physical world and the top-down interpretation of disentanglement of the holistic representation conjointly shape our subjective experience in terms of temporal sequence. The current study is the first to provide behavioral evidence that there exists a dissociation between the two phases of temporal order reconstruction, and more neural and modeling evidence can be further investigated by future studies.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Projects 31771213 and 31371018). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Zhilin Zhang https://orcid.org/0000-0002-0847-891X
Xiaoxiong Lin https://orcid.org/0000-0002-6212-0249
Yan Bao https://orcid.org/0000-0002-5907-3955

REFERENCES

Babkoff, H., & Fostick, L. (2013). The role of tone duration in dichotic temporal order judgment. *Attention, Perception, & Psychophysics*, *75*(4), 654–660. https://doi.org/10.3758/s13414-013-0449-6

Bao, Y., Fang, Y., Yang, T., Wang, L., Szymaszek, A., & Szelag, E. (2014). Auditory perception of temporal order: A comparison between tonal language speakers with and without non-tonal language experience. *Acta Neurobiologiae Experimentalis*, *74*(1), 98–103.

Bao, Y., Pöppel, E., Wang, L., Lin, X., Yang, T., Avram, M., … Zaytseva, Y. (2015). Synchronization as a biological, psychological and social mechanism to create common time: A theoretical frame and a single case study. *PsyCh Journal*, *4*(4), 243–254. https://doi.org/10.1002/pchj.119

Bao, Y., Szymaszek, A., Wang, X., Oron, A., Pöppel, E., & Szelag, E. (2013). Temporal order perception of auditory stimuli is selectively modified by tonal and non-tonal language environments. *Cognition*, *129*(3), 579–585. https://doi.org/10.1016/j.cognition.2013.08.019

Bao, Y., von Stosch, A., Park, M., & Pöppel, E. (2017). Complementarity as generative principle: A thought pattern for aesthetic appreciations and cognitive appraisals in general. *Frontiers in Psychology*, *8*, 727. https://doi.org/10.3389/fpsyg.2017.00727

Bennett, R. G., & Westheimer, G. (1985). A shift in the perceived simultaneity of adjacent visual stimuli following adaptation to stroboscopic motion along the same axis. *Vision Research*, *25*(4), 565–569. https://doi.org/10.1016/0042-6989(85)90161-0

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897x00357

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*(10), 755–765. https://doi.org/10.1038/nrn1764

Cai, M., Stetson, C., & Eagleman, D. M. (2012). A neural model for temporal order judgments and their active recalibration: A common mechanism for space and time? *Frontiers in Psychology*, *3*, 470. https://doi.org/10.3389/fpsyg.2012.00470

Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science*, *12*(6), 532–535. 10.1111%2F1467-9280.d01-17.

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12), 7. https://doi.org/10.1167/9.12.7

Fekete, T., Van de Cruys, S., Ekroll, V., & van Leeuwen, C. (2018). In the interest of saving time: A critique of discrete perception. *Neuroscience of Consciousness*, *2018*(1), niy003. https://doi.org/10.1093/nc/niy003

Fingelkurts, A. A., & Fingelkurts, A. A. (2006). Timing in cognition and EEG brain dynamics: Discreteness versus continuity. *Cognitive Processing*, *7*(3), 135–162. https://doi.org/10.1007/s10339-006-0035-0

Fink, M., Ulbrich, P., Churan, J., & Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behavioural Processes*, *71*(2–3), 344–352. https://doi.org/10.1016/j.beproc.2005.12.007

Fostick, L., & Babkoff, H. (2013). Different response patterns between auditory spectral and spatial temporal order judgment (TOJ). *Experimental Psychology*, *60*(6), 432–443. https://doi.org/10.1027/1618-3169/a000216

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. Y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

Fuzessery, Z. M., Richardson, M. D., & Coburn, M. S. (2006). Neural mechanisms underlying selectivity for the rate and direction of frequency-modulated sweeps in the inferior colliculus of the pallid bat. *Journal of Neurophysiology*, *96*(3), 1320–1336. https://doi.org/10.1152/jn.00021.2006

Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: Long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, *24*(10), 826–837. https://doi.org/10.1016/j.tics.2020.07.001

Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*(5), 423–432. https://psycnet.apa.org/doi/10.1037/h0045283

Holcombe, A. (2015). The temporal organization of perception. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 820–840). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199686858.013.057

Kanabus, M., Szelag, E., Rojek, E., & Pöppel, E. (2002). Temporal order judgement for auditory and visual stimuli. *Acta Neurobiologiae Experimentalis*, *62*(4), 263–270.

Kleiner, M., Brainard, D., & Pelli, D., 2007, "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement.

Koenderink, J., Richards, W., & van Doorn, A. J. (2012). Space-time disarray and visual awareness. *Perception*, *3*(3), 159–165. 10.1068%2Fi0490sas.

Lee, G. (2014). Temporal experience and the temporal structure of experience. *Philosophers' Imprint*, *14*, 1–21.

Liang, W., Zhang, J., & Bao, Y. (2015). Gender-specific effects of emotional modulation on visual temporal order thresholds. *Cognitive Processing*, *16*(1), 143–148. https://doi.org/10.1007/s10339-015-0709-6

Miyazaki, M., Kadota, H., Matsuzaki, K. S., Takeuchi, S., Sekiguchi, H., Aoyama, T., & Kochiyama, T. (2016). Dissociating the neural correlates of tactile temporal order and simultaneity judgements. *Scientific Reports*, *6*, 23323. 10.1038%2Fsrep23323.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, *9*(7), 875–877. https://psycnet.apa.org/doi/10.1038/nn1712

Notter, M. P., Hanke, M., Murray, M. M., & Geiser, E. (2019). Encoding of auditory temporal gestalt in the human brain. *Cerebral Cortex*, *29*(2), 475–484. https://doi.org/10.1093/cercor/bhx328

Okada, M., & Kashino, M. (2003). The role of spectral change detectors in temporal order judgment of tones. *Neuroreport*, *14*(2), 261–264. https://psycnet.apa.org/doi/10.1097/00001756-200302100-00021

Pöppel, E. (1970). Excitability cycles in central intermittency. *Psychologische Forschung*, *34*(1), 1–9. https://doi.org/10.1007/bf00422860

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*(2), 56–61. https://doi.org/10.1016/S1364-6613(97)01008-5

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1887–1896. https://doi.org/10.1098/rstb.2009.0015

Pöppel, E., & Bao, Y. (2014). Temporal windows as a bridge from objective to subjective time. In D. Lloyd & V. Arstila (Eds.), *Subjective time* (pp. 241–261). MIT Press.

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1710), 1314–1322. https://doi.org/10.1098/rspb.2010.1737

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, *51*(5), 651–659. https://doi.org/10.1016/j.neuron.2006.08.006

Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, *116*(2), 918–933. https://doi.org/10.1121/1.1763601

Takahashi, T., Kansaku, K., Wada, M., Shibuya, S., & Kitazawa, S. (2013). Neural correlates of tactile temporal-order judgment in humans: An fMRI study. *Cerebral Cortex*, *23*(8), 1952–1964. https://doi.org/10.1093/cercor/bhs179

Tian, B., & Rauschecker, J. P. (2004). Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, *92*(5), 2993–3013. https://doi.org/10.1152/jn.00472.2003

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception & Psychophysics*, *42*(3), 224–239. https://doi.org/10.3758/bf03203074

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213. https://doi.org/10.1016/S1364-6613(03)00095-0

White, P. A. (2018). Is conscious perception a series of discrete temporal frames? *Consciousness and Cognition*, *60*, 98–126. https://doi.org/10.1016/j.concog.2018.02.012

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://psycnet.apa.org/doi/10.3758/BF03194544

Zhou, B., Pöppel, E., & Bao, Y. (2014). In the jungle of time: The concept of identity as a way out. *Frontiers in Psychology*, *5*, 844. https://doi.org/10.3389/fpsyg.2014.00844

Zhuo, Y., Zhou, T. G., Rao, H. Y., Wang, J. J., Meng, M., Chen, M., Zhou, C., & Chen, L. (2003). Contributions of the visual ventral pathway to long-range apparent motion. *Science*, *299*(5605), 417–420. https://doi.org/10.1126/science.1077091

How to cite this article: Zhang, Z., Lin, X., & Bao, Y. (2023). Holistic temporal order judgment of tones requires top-down disentanglement. *PsyCh Journal*, *12*(4), 491–499. https://doi.org/10.1002/pchj.603