Cognition 129 (2013) 579–585

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Temporal order perception of auditory stimuli is selectively modified by tonal and non-tonal language environments

Yan Bao [a,b,*,1], Aneta Szymaszek [c], Xiaoying Wang [a], Anna Oron [c], Ernst Pöppel [a,b], Elzbieta Szelag [b,c,1]

[a] *Department of Psychology & Key Laboratory of Machine Perception (MoE), Peking University, Beijing 100871, PR China*
[b] *Human Science Center & Institute of Medical Psychology, Ludwig Maximilian University Munich, 80336 München, Germany*
[c] *Laboratory of Neuropsychology, Nencki Institute of Experimental Biology, 02-093 Warsaw, Poland*

ARTICLE INFO

Article history:
Received 9 December 2011
Revised 15 August 2013
Accepted 19 August 2013
Available online 21 September 2013

Keywords:
Time perception
Auditory processing
Temporal order threshold
Language experience
Double dissociation of function

ABSTRACT

The close relationship between temporal perception and speech processing is well established. The present study focused on the specific question whether the speech environment could influence temporal order perception in subjects whose language backgrounds are distinctively different, i.e., Chinese (tonal language) vs. Polish (non-tonal language). Temporal order thresholds were measured for both monaurally presented clicks and binaurally presented tone pairs. Whereas the click experiment showed similar order thresholds for the two language groups, the experiment with tone pairs resulted in different observations: while Chinese demonstrated better performance in discriminating the temporal order of two "*close frequency*" tone pairs (600 Hz and 1200 Hz), Polish subjects showed a reversed pattern, i.e., better performance for "*distant frequency*" tone pairs (400 Hz and 3000 Hz). These results indicate on the one hand a common temporal mechanism for perceiving the order of two monaurally presented stimuli, and on the other hand neuronal plasticity for perceiving the order of frequency-related auditory stimuli. We conclude that the auditory brain is modified with respect to temporal processing by long-term exposure to a tonal or a non-tonal language. As a consequence of such an exposure different cognitive modes of operation (analytic vs. holistic) are selected: the analytic mode is adopted for "*distant frequency*" tone pairs in Chinese and for "*close frequency*" tone pairs in Polish subjects, whereas the holistic mode is selected for "*close frequency*" tone pairs in Chinese and for "*distant frequency*" tone pairs in Polish subjects, reflecting a double dissociation of function.

1. Introduction

Temporal information processing plays an important role in cognitive processes like perception, attention, working memory, movement control or language (for an overview, see Fingelkurts & Fingelkurts, 2006). One domain of temporal processing is related to sequential timing in the range of some tens of milliseconds (Fink, Churan, & Wittmann, 2006; Pöppel, 2009; Szymaszek, Sereda, Pöppel, & Szelag, 2009). In order to discriminate accurately the temporal order of two successively presented stimuli, independent of stimulus modality (auditory, visual, or tactile) a

---

* Corresponding author. Address: Department of Psychology, Peking University, 5 Yiheyuan Road, Beijing 100871, PR China. Tel.: +86 10 62753200; fax: +86 10 62761081.

E-mail addresses: baoyan@pku.edu.cn (Y. Bao), a.szymaszek@nencki.gov.pl (A. Szymaszek), wangxiaoy@pku.edu.cn (X. Wang), a.oron@nencki.gov.pl (A. Oron), Ernst.Poeppel@med.uni-muenchen.de (E. Pöppel), szelag@nencki.gov.pl (E. Szelag).

[1] These authors contributed equally to this work.

0010-0277
http://dx.doi.org/10.1016/j.cognition.2013.08.019

time interval of approximately the same duration is required between the onsets of the two stimuli (Hirsh & Sherrick, 1961). This minimum time interval is termed as *temporal order threshold* (*TOT*), which indicates a necessary temporal interval for establishing the before–after relationship of successive stimuli. A lower TOT indicates better temporal order processing.

Previous studies have revealed a number of factors that can influence TOT such as age, gender, stimulus type or presentation mode. Young children usually have difficulties to perform the temporal order task and their TOTs tend to decrease as they grow older (Berwanger, Wittmann, von Steinbüchel, & von Suchodoletz, 2004). Compared to young adults, elderly people usually demonstrate higher TOTs (Kolodziejczyk & Szelag, 2008; Szymaszek, Szelag, & Sliwowska, 2006; Ulbrich, Churan, Fink, & Wittmann, 2009). Men tend to have lower TOT than women for discriminating two identical clicks, each being presented to one ear (Lotze, Wittmann, von Steinbuchel, Pöppel, & Roenneberg, 1999; Wittmann & Szelag, 2003). Compared to monaurally presented clicks, binaurally presented tones usually result in lower TOT as observed in various subjects groups (Fink, Ulbrich, Churan, & Wittmann, 2006).

Although several factors influencing TOT have been addressed, no sufficient evidence is available regarding whether speech experience itself can influence TOT. This question is important since a close link between temporal information processing and language capabilities has been suggested in previous research. Children with language learning impairment and patients with acquired aphasia or dyslexia often demonstrate both language disabilities (e.g., deficits in phoneme identification and/or discrimination) and deteriorated timing (e.g., increased TOT for detecting the temporal order of sequentially presented acoustic stimuli) (Ben-Artzi, Fostick, & Babkoff, 2005; Gaab, Gabrieli, Deutsch, Tallal, & Temple, 2007; Sidiropoulos, Ackermann, Wannke, & Hertrich, 2010; Szelag, von Steinbüchel, & Pöppel, 1997; Tallal, Merzenich, Miller, & Jenkins, 1998; Tallal et al., 1996; Vandermosten et al., 2011; Wittmann & Fink, 2004). Based on these observations temporal order processing as an underlying basis for language comprehension is assumed. However, a reversed consideration might also be interesting, i.e., whether language experience may influence temporal order processing.

Apparently, only one study thus far addressed this question (Szelag et al., 2011). The results seem to suggest a common and language-independent mechanism for temporal order processing, since no language effect on TOT was observed. However, not having observed such an effect on TOT might have had several reasons. First, the subjects tested in their study were German and Polish, whose native languages are different in many aspects, but also share important similarities, since both German and Polish fall into the same category of non-tonal languages. Thus, these two languages might be not sufficiently different for capturing a potential language effect on temporal order processing. Second, the acoustic stimuli used for measuring TOT in their study were two identical brief stimuli (1 ms clicks) which are qualitatively different from two stimuli such as tones with different frequencies. Successive presentation of two tones with different frequencies mimics better the frequency variations that one experiences in a natural speech environment. Third, the two clicks in their temporal order measurements were presented monaurally with one click to one ear (e.g., left ear) and the other click to the other ear (e.g., right ear). This stimulus presentation mode is obviously different from a normal auditory environment, in which speech signals are simultaneously received from both ears.

In order to answer the question whether language experience influences temporal order processing the temporal characteristics of tonal and non-tonal languages have to be addressed in more detail. It is well known that temporal cues in speech signals play an important role in decoding syllables or words into their phonemic segments for auditory comprehension. For example, to discriminate "duck" and "tuck" in English, or "dui" and "tui" in Chinese, the *voice-onset-time* (VOT, the time distance between the burst and the onset of laryngeal pulsing) for voiced and unvoiced consonants (/d/ and /t/) plays a crucial role, since the VOT in many languages has different duration. Besides this general aspect, tonal and non-tonal languages differentiate themselves in some temporal characteristics such as the number of syllables in a word, the duration of consonants or vowels. A tonal language such as Chinese is mainly monosyllabic, while non-tonal languages such as English, German and Polish are typically multisyllabic. The duration of vowels in Chinese generally lasts longer than that in non-tonal languages. However, the most salient difference in temporal characteristics between tonal and non-tonal languages is related to pitch contour (Kann, Wayland, Bao, & Barkley, 2007; Krishnan, Gandour, & Bidelman, 2010; Luo, Boemio, Gordon, & Poeppel, 2007). Unlike non-tonal languages such as German and Polish, the meaning of a word in a tonal language such as Chinese cannot be solely defined by consonants and vowels without a lexical tone. For example, the Chinese syllable /ba/ may has four distinct lexical meanings when spoken with different pitch contours. It can mean the digit "8" when pronounced with a *high level* tone, the action "pulling up" with a *high rising* tone, the "target" for hunting with a *low dipping* tone, or the appellation "father" with a *high falling* tone. Therefore, to extract the meaning of Chinese words, the pitch contour which features small changes in frequency range plays a crucial role. In contrast, non-tonal languages such as English, Germany and Polish only have one single lexical meaning in one syllable regardless of some possible tone variations; thus, it is the pitch height, not the pitch contour, which is important in decoding semantic information. In other words, non-tonal languages are characterized by the large changes in frequency range, and the pitch contour plays barely any role in decoding lexical meanings.

Considering the limitations in the study mentioned above (Szelag et al., 2011) and the major difference between tonal and non-tonal languages as outlined above, the question of whether a language environment may influence temporal order processing has to be addressed again. Therefore, we designed two experiments in the present study. Experiment 1 aimed to replicate the study by Szelag et al. (2011) using monaurally presented clicks in two other language groups, i.e. Chinese and Polish subjects. The purpose of this experiment was to test whether the previous finding could also be observed in subjects whose language environments were distinctively different. If a common temporal mechanism exists, then no TOT difference should be observed between the two language groups. If there is, however, a language impact on temporal order processing, and if this effect can even be captured by monaurally presented click measurements, then significantly different TOTs should be observed for the two language groups.

In Experiment 2 we used binaurally presented two tones of different frequencies as stimuli to measure the TOTs. In order to capture the most salient difference between tonal and non-tonal languages, we manipulated the frequency distance between the two tones and tested two types of tone pairs. One is "*close frequency*" tone pairs using 600 Hz and 1200 Hz tones as stimuli; the other is "*distant frequency*" tone pairs using 400 Hz and 3000 Hz tones. Since tonal language subjects such as Chinese are experienced in detecting pitch contours in their natural language environment, we assume that for Chinese subjects a lower TOT for the "*close frequency*" relative to the "*distant frequency*" tones might be observed, since for Chinese subjects the temporal order task using two "*close frequency*" tones might be treated as a task of detecting *pitch contours* with either an *up* or a *down* patterning of frequency steps. Since Polish subjects rarely experience detecting pitch contours in their natural speech environment, equally high TOTs are anticipated for discriminating the temporal order of both the "*close frequency*" and the "*distant frequency*" tones.

2. Methods

2.1. Participants

Two groups of volunteers with different language background (*NP group*: 18 native Polish speakers; *NC group*: 18 native Chinese Mandarin speakers) participated in the present study. Each language group was half male and half female, aged on average 25 years. The NP participants were recruited in the area of Warsaw and had no previous experience of Chinese. The NC participants were recruited in the area of Beijing and had no previous experience of Polish. All participants were right-handed (Oldfield, 1971) and had no history of neurological or psychiatric disorders, or any indication of cognitive impairment. All participants had received a college education with years varied slightly due to the different education systems in each country. None of the participants had received professional musical education in special schools or universities, which possibly could increase the sensitivity of auditory perception of acoustic stimuli.

To ensure normal hearing in all participants, pure-tone audiometry screening was performed (Audiometer AS 208 or GSI 17). The adaptive procedure used frequencies ranging from 250 Hz to 3000 Hz (250, 500, 750, 1000, 1500, 2000, 3000 Hz) and a dB range from −10 dB to 100 dB in steps of 5 dB. The criteria for admission to the present study were a hearing level below 30 dB for all frequencies tested and differences in hearing level between the two ears below 20 dB (ANSI, 2004).

To assess the intellectual abilities of the two language groups, non-verbal Mosaic Test from the widely used HAWIE-R test battery (see Tewes, 1994) was performed. This test is the subtest of the HAWIE-R correlating highest with the total score of the non-verbal activity part of the HAWIE-R, thus being an indicator of general fluid intelligence. During the Mosaic Test subjects were asked to copy a pattern of red and white squares presented on a picture using four or nine red and white cubes. A total number of nine different patterns were presented. The first five patterns were easier requiring only four cubes. The last four patterns were more difficult and had to be copied with nine cubes. Subjects would receive scores when they replicated the patterns correctly within a given time frame of 60 s (four cubes) or 120 s (nine cubes). The faster they solved each pattern, the higher the score they would receive. For both the NP and the NC participants, matched Mosaic scores were obtained. Detailed descriptive data of the participants are listed in Table 1.

Separate studies in Warsaw and in Beijing for the two language groups were conducted using identical experimental procedures. Each participant was tested individually in two separate sessions in two different days in a soundproof room. Informed consent was provided before and rewards were provided afterward.

2.2. Materials and procedure

All stimuli used for temporal order measurements were pairs of acoustic stimuli, which were generated with the program Cool Edit 2000 (sampling rate 44,100 Hz, 16-bit). The two paired stimuli were presented in a rapid succession with varied inter-stimulus intervals (ISIs) which were controlled by a YAAP (Treutwein, 1997) algorithm. Details are described below.

2.2.1. Experiment 1

Experiment 1 used pairs of clicks as stimuli. The clicks were 1 ms rectangular pulses presented in an alternating monaural stimulation mode, i.e., one click was presented to one ear, followed by another click to the other ear. Subjects were asked to indicate the sequence of the two clicks by pointing to one of the two response cards: "left–right" or "right–left".

2.2.2. Experiment 2

Experiment 2 used two types of paired sinusoidal tones as stimuli. One was "*close frequency*" tones consisting of a low tone of 600 Hz and a high tone of 1200 Hz. The other was "*distant frequency*" tones consisting of a low tone of 400 Hz and a high tone of 3000 Hz. The duration of each tone was 10 ms with 1 ms rise-and-fall time. The two tones in each trial were presented in a binaural stimulation mode, i.e., each tone was presented to both ears with a short gap in between. The subjects had to indicate the temporal order of the two tones by pointing to one of the two response cards: "low–high" or "high-low".

Table 1
Characteristics of participants.

Group	*N* (male/female)	Age range	Mean age (SD)	Handedness (left or right)	Mosaic (SD)	Hearing status (normal or not)
NC	18 (9/9)	21–29	25 (2.4)	Right	37 (7.3)	Normal
NP	18 (9/9)	20–28	25 (2.3)	Right	37 (7.1)	Normal

Abbreviations: NC = Native Chinese speakers; NP = Native Polish speakers; *N* = Number of participants; SD = Standard Deviation.

2.2.3. YAAP procedure

In both experiments, the inter-stimulus intervals between the two acoustic stimuli were controlled by a maximum-likelihood based algorithm – YAAP (Treutwein, 1997) procedure. According to the subjects' previous responses, the ISI of the present trial was set at the current best estimate of the threshold corresponding to 75% correct responses based on a logistic psychometric function. The YAAP test included two phases. In an initial phase, ten pairs of stimuli were presented. The first inter-stimulus interval started at 80% of a pre-defined upper limit (200 ms in both Exp1 and Exp2) and proceeded in equal steps of 20% to a specified lower limit (10 ms in Exp1 and 1 ms in Exp2, yielding a same minimum SOA of 11 ms). In the second phase, presented inter-stimulus intervals were based on the estimation process of the YAAP algorithm. The stimulus presentation was terminated when the location of the threshold parameter was with a probability of 95% inside a ±5 ms interval around the currently estimated threshold (Treutwein, 1995).

For all measurements, each trial started with three warning signals prior to the presentation of the paired stimuli to focus the subjects' attention on the coming task. The time between the warning signal and the first stimulus was 1500 ms. All stimuli were presented via a headphone (SONY MDR-CD 480) at a comfortable listening level which is well above threshold. The frequency discrimination thresholds for the four tones used (400 Hz, 600 Hz, 1200 Hz, and 3000 Hz) were not measured, since the frequency distance in all tone pairs were far beyond the bandwidth or difference limens of the four frequencies selected (Zwicker, Flottorp, & Stevens, 1957). The main measurement in each experiment was preceded by a practice session in which participants reported the temporal order of the two acoustic stimuli presented with a constant, relatively long ISI of 160 ms. The practice was continued until a criterion of 11 correct responses in a series of consecutive 12 presentations was reached.

3. Results

3.1. Temporal order threshold (TOT)

The sequencing ability was indexed by the *temporal order threshold* (TOT) which was defined as the minimum time interval required for correctly identifying the temporal order of two successively presented stimuli. According to previous studies, either the *inter-stimulus interval* (ISI), i.e., the time interval between the offset of the first stimulus and the onset of the second stimulus, or the *stimulus onset asynchrony* (SOA), i.e., the time interval between the onset of the first stimulus and the onset of the second stimulus, can be used as index of TOT. Since the presentation time of clicks (1 ms) in Experiment 1 and the presentation time of tones (10 ms) in Experiment 2 were different, we chose to use SOA (i.e., ISI + duration of the first stimulus) to measure the TOT in both experiments so as to have a unified index for comparison.

To obtain individual values of TOT, a logistic psychometric function was fitted to the subject's data, using MATLAB toolbox psignifit version 2.5.41 (see http://bootstrapsoftware.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann and Hill (2001). This procedure estimates an ISI corresponding to 75% correct order discrimination (Strasburger, 2001). We further calculated SOA by adding the stimulus duration (clicks: 1 ms; tones: 10 ms) to the ISI to obtain the TOT values for later statistical analysis. For example, an ISI of 20 ms represents a SOA of 21 ms for click measurements and 30 ms for tone measurements. In both experiments, an averaged TOT value of two separate sessions for each subject was calculated and used as dependent variable for later statistical analysis in the present study.

3.2. Experiment 1

For click experiment a two-way ANOVA was performed with subject *Group* (Chinese vs. Polish) and *Gender* (male vs. female) as two independent variables. The results demonstrated a significant main effect of *Gender*, $F_{(1,32)} = 4.224$, $P < 0.05$, $\eta_p^2 = 0.117$. Males showed a lower TOT than females, confirming a better temporal processing of monaurally presented clicks in males (Lotze et al., 1999; Wittmann & Szelag, 2003). The main effect of subject *Group*

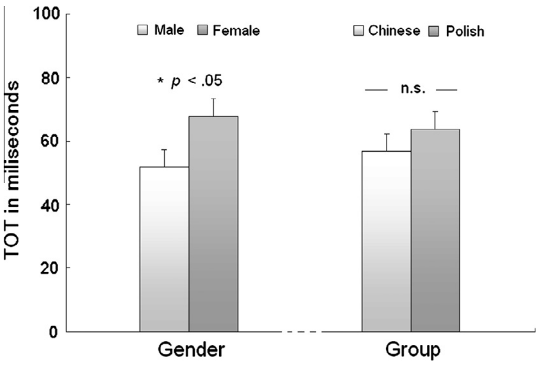

Fig. 1. Temporal order thresholds (TOTs) measured in Experiment 1 using click stimuli. The left panel illustrates a significant main effect of gender which did not interact with any other variables. The right panel illustrates the non-significant main effect of subject group (Chinese vs. Polish).

was not significant, $F_{(1,32)} = 0.919$, $P = 0.345$, $\eta_p^2 = 0.028$. The two-way interaction was not significant either, $F_{(1,32)} = 0.127$, $P = 0.724$, $\eta_p^2 = 0.004$. See Fig. 1 for the visualized main effects of *Gender* and subject *Group*. The results of Experiment 1 indicate that temporal order processing as measured with monaurally presented clicks is independent of subjects' tonal and non-tonal language backgrounds, which is consistent with previous observation (Szelag et al., 2011).

3.3. Experiment 2

For tone experiment a three-way mixed ANOVA was performed with subject *Group* (Chinese vs. Polish) and *Gender* (male vs. female) as two between-subjects variables and the stimulus *Type* (*close frequency* vs. *distant frequency* tone pairs) as one within-subjects variable. All main effects and interactions were non-significant except the two-way interaction between stimulus *Type* and subject *Group*, $F_{(1,32)} = 26.593$, $P < 0.001$, $\eta_p^2 = 0.454$. This highly significant two-way interaction is illustrated in Fig. 2, and it has to be emphasized that this interaction was the same for both male and female subjects since no gender effect and no interaction between gender and other variables were observed. Further analysis of this interaction revealed very interesting TOT patterns between the two language groups: For the *NC* group, a significantly lower TOT was observed for discriminating the temporal order of two *close frequency* tones (600 Hz and 1200 Hz) relative to the two *distant frequency* tones (400 Hz and 3000 Hz) (27 ms vs. 51 ms, $P < 0.01$). However, for the *NP* group, a reversed pattern was observed, i.e., a significantly higher TOT was demonstrated for the *close frequency* tone pairs as compared to the *distant frequency* tone pairs (64 ms vs. 30 ms, $P < 0.001$). This double dissociation result pattern indicates an important language impact on temporal order processing of binaurally presented tone stimuli.

4. Discussion

The present study measured temporal order thresholds with both monaurally presented clicks (Exp1) and binaurally presented tones (Exp2) in both tonal (Chinese) and non-tonal (Polish) language speakers. The results of Exp1 confirmed the previous observation by Szelag et al. (2011). Like in the comparison of Polish and German subjects no TOT difference between Chinese and Polish subjects was observed. Thus, the click TOT might indeed tap a general phenomenon in temporal processing, i.e., a common neural mechanism underlying milliseconds timing which governs click ordering. This mechanism may set a fundamental basis of temporal processing for both tonal and non-tonal language speakers. In fact, the observed gender effect (i.e., males performed better than females in click TOT measurement, which was consistent with previous findings) which did not interact with any other factors further indicated that this common temporal mechanism exists similarly for males and females in both tonal and non-tonal language environments.

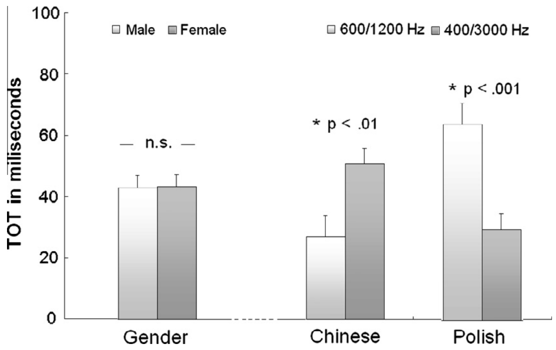

Fig. 2. Temporal order thresholds (TOTs) measured in Experiment 2 using two types of tone stimuli. The left panel illustrates the non-significant gender effect, and the right panel illustrates the significant interaction between stimulus type and subject group (a double dissociation TOT pattern), which is independent of gender.

However, as predicted in the introduction, a language impact on tone order thresholds was captured in Exp2. For Chinese subjects who were experienced with detection of pitch contours, a significantly lower TOT was observed for "*close frequency*" tones relative to "*distant frequency*" tones. This difference was not shown for Polish subjects whose language is not characterized by pitch contours. It came as a surprise, however, that the Polish subjects showed also a TOT difference, i.e., a lower order threshold for the "*distant frequency*" tones as compared to the "*close frequency*" tones. Thus, a double dissociation was observed which was not predicted. Moreover, different from click experiment (Exp1), no gender effect was observed in the tone experiment (Exp2). The fact that in one case it was observed and in the other it was not further supports the idea that different temporal mechanisms are involved in click and tone order processing.

As already known, subjects might use different processing modes or "strategies" to decode the sequence of the two acoustic events. One such strategy can be referred to as "*analytic*": Subjects have to identify the singular acoustic events in their identity and on that basis they have a direct experience of the temporal order of these events. For example, in the click experiment the two clicks were presented monaurally, i.e., each being presented to one ear. Thus, the click experiment consists not only of a temporal order task, but also a spatial localization task. Therefore, the subjects may localize the source of the first click (*left* or *right*) and then the source of the second click (*right* or *left*) immediately after the stimulus presentation, and the temporal order judgment is directly based on such spatial information of the stimuli. Similarly, in the tone experiment although the two acoustic stimuli were presented binaurally (thus no spatial information was available), they were different in frequencies; one is a *high* tone, and the other is a *low* tone. Thus, the tone experiment involves not only a temporal order task, but also a frequency discrimination task. Therefore, the subjects may identify the frequency of each tone immediately following the presentation of the stimulus, and on that basis report the temporal order of the two tones.

Another strategy to build up a temporal relationship of two acoustic stimuli can be referred to as "holistic": The two successively presented acoustic events are integrated into one unitary percept with either a *rising* or a *falling* frequency modulated sweep. On the basis of recognizing this pitch contour like global patterning (frequency goes *up* or goes *down*), subjects reconstruct secondarily the temporal order of the two stimuli (reporting "high–low" or "low–high"), leaving the identification of each individual tone unnecessary (Brechmann & Scheich, 2005; Szymaszek et al., 2009). Such a holistic processing must be based on acoustic information starting from the onset of the first stimulus until the end of the second stimulus in order for subjects to perceive a global frequency variation pattern. Compared to the analytic processing, the holistic processing usually leads to a lower temporal order threshold.

The results of the present study can be interpreted by the two different operative modes of temporal order processing. In the click experiment where each click was presented for 1 ms with a varied interval, it was impossible to integrate the two clicks into one unitary percept when they are separated by a gap exceeding 3–5 ms (Fink, Churan, et al., 2006). Therefore, an analytic mode was adopted for reporting the temporal order of two clicks. In the tone experiment, when the time interval between the two successive tones was relatively long, a similar analytic processing mode was presumably adopted for reporting the temporal order of the two tones. When the time interval decreased, a holistic processing mode might operate in a way that subjects integrated the two tones into one percept with either an *up* (from low frequency to high frequency) or a *down* (from high frequency to low frequency) spectral glide, and the temporal order of the two tones was reconstructed secondarily from the direction of the frequency modulated sweeps.

Interestingly as revealed by the tone experiment, such a holistic operating mode was apparently modified by long-term exposure to a native language environment. As predicted, the Chinese subjects seemed to use a holistic processing for the two "*close frequency*" tones, leading to a significantly lower TOT than for the two "*distant frequency*" tones. However, the Polish subjects, who were not experienced with "pitch contour" detection in their natural language environment, also showed surprisingly a lower TOT but for the "*distant frequency*" tones. Thus, one has to argue that Polish subjects also use *holistic* processing for detecting the temporal order of the two "*distant frequency*" tones. This implies that Polish subjects have a higher sensitivity to large frequency variations possibly due to the phonetic characteristics of their non-tonal language. For Polish subjects, the "*close frequency*" tones do not elicit a global processing presumably due to the fact that their non-tonal language environment has not produced a specific sensitivity for small variations of frequency change as compared to Chinese subjects. Thus, they adopt an analytic strategy for perceiving the temporal order of the two "*close frequency*" tones.

Our interpretation of the holistic processing mode being shaped by tonal or non-tonal language environment seems to be consistent with previous single neuron recordings, which suggest that neurons in the primary auditory cortex are selective to directions of frequency modulated sweeps (e.g., Tian & Rauschecker, 2004). Possibly, such neurons are differentially stimulated during the long-term exposure to a tonal or a non-tonal language environment, leading to different sensitivities of such neurons. Thus, a flexible holistic processing tuned by native language environment, perhaps, indicates a neural plasticity as suggested in a previous study (Merzenich & Sameshima, 1993).

The results of our tone experiment (Exp2) support the hypothesis that language environment has an influence on temporal processing. Thus, one has to conclude that not only temporal processing affects speech processing as referred to in the introduction, but that the reverse is also true. This leads to a logical challenge, i.e., temporal processing underlying speech processing is itself altered by speech processing. This is obviously a circular argument. How does the brain get out of this bi-directional interaction? One may argue that such an impact of language on sequential timing works only within temporal limits.

There is evidence for a "temporal window" of some tens of milliseconds as observed with a number of different experimental paradigms which suggest the general principle of discrete time sampling in neuronal information processing, and on that basis indicating physiological temporal limits which for integration of neuronal information presumably cannot be transcended. Such an interval of temporal integration has to be assumed on the basis of a theoretical account: information to be processed is not available at pre-defined time points because of different transduction time in various sensory modalities (e.g., being much shorter in the auditory compared to the visual modality); furthermore, the physical distance of stimuli is relevant in the auditory but not in the visual modality under ecological conditions resulting in a particular challenge for intersensory integration. Thus, temporal windows have been proposed as system states for integration of neuronal information (Pöppel, 1997, 2009; Pöppel, Schill, & von Steinbüchel, 1990).

Empirical support for this concept of discrete time sampling in the temporal range of some tens of milliseconds comes for instance from the multimodal response distributions as measured by choice reaction time (Pöppel, 1970), pursuit eye movement latencies (Pöppel & Logothetis, 1986), and anticipatory movements during sensorimotor synchronization (Radil, Mates, Ilmberger, & Pöppel, 1990). Similarly, oscillations of neuronal populations with periods in the domain of some tens of milliseconds (Galambos, Makeig, & Talmachoff, 1981; Madler & Pöppel, 1987) as well as single cell activities as observed in the accessory visual pathway (Podvigin, Jokeit, Pöppel, Chizh, & Kiselyeva, 1992; Podvigin et al., 2004) can also be conceived of being expressions of discrete time sampling in the same temporal window. Furthermore, evidence on attentional control in the visual field suggests a stronger inhibitory function in the more peripheral relative to the perifoveal visual field indexed by a delayed response time by some tens of milliseconds corresponding to the numerical values observed in other experimental paradigms mentioned above (Bao & Pöppel, 2007). Taken together, a rather stable "temporal window" in cognitive processing is suggested which may provide a temporal frame which cannot be transcended in adaptive processes. Thus, the circular argument that temporal processing affects speech processing and speech processing affects temporal processing may only be true within "temporal windows" of neuronal processing that define temporal limits.

Acknowledgements

This research was supported by the National Nature Science Foundation of China (Projects 30670703 and 91120004) and the Polish Ministry of Science and Higher Education (No. 507/1/N-DFG/2009/0).

References

ANSI (2004). *ANSI S3.6-2004. American national standard specification for audiometers*. New York: American National Standards Institute.

Bao, Y., & Pöppel, E. (2007). Two spatially separated attention systems in the visual field: Evidence from inhibition of return. *Cognitive Processing, 8*, 37–44.

Ben-Artzi, E., Fostick, L., & Babkoff, H. (2005). Deficits in temporal-order judgments in dyslexia: Evidence from diotic stimuli differing spectrally and from dichotic stimuli differing only by perceived location. *Neuropsychologia, 43*, 714–723.

Berwanger, D., Wittmann, M., von Steinbüchel, N., & von Suchodoletz, W. (2004). Measurement of temporal-order judgment in children. *Acta Neurobiologiae Experimentalis, 64*, 387–394.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sounds representation in auditory cortex with conceptual learning. *Cerebral Cortex, 15*, 578–587.

Fingelkurts, A. A., & Fingelkurts, A. A. (2006). Timing in cognition and EEG brain dynamics: Discreteness versus continuity. *Cognitive Processing, 7*, 135–162.

Fink, M., Churan, J., & Wittmann, M. (2006). Assessment of auditory temporal order thresholds – A comparison of different measurement procedures and the influence of age and gender. *Restorative Neurology and Neuroscience, 23*, 1–16.

Fink, M., Ulbrich, P., Churan, J., & Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behavioral Processes, 71*, 344–352.

Gaab, N., Gabrieli, J. D., Deutsch, G. K., Tallal, P., & Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: An fMRI study. *Restorative Neurology and Neuroscience, 25*, 295–310.

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences of the United States of America, 78*, 2643–2647.

Hirsh, I. J., & Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62*, 423–432.

Kann, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research, 1148*, 113–122.

Kolodziejczyk, I., & Szelag, E. (2008). Auditory perception of temporal order in centenarians in comparison with young and elderly subjects. *Acta Neurobiologiae Experimentalis, 68*, 373–381.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics, 23*, 81–95.

Lotze, M., Wittmann, M., von Steinbuchel, N., Pöppel, E., & Roenneberg, T. (1999). Daily rhythm of temporal resolution in the auditory system. *Cortex, 35*, 89–100.

Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007). The perception of FM sweeps by Chinese and English listeners. *Hearing Research, 224*, 75–83.

Madler, Ch., & Pöppel, E. (1987). Auditory evoked potentials indicate the loss of neuronal oscillations during general anaesthesia. *Naturwissenschaften, 74*, 42–43.

Merzenich, M. M., & Sameshima, K. (1993). Cortical plasticity and memory. *Current Opinion in Neurobiology, 3*, 187–196.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Podvigin, N. F., Bagaeva, T. V., Boykova, E. V., Zargarov, A. A., Podvigina, D. N., & Pöppel, E. (2004). Three bands of oscillatory activity in the lateral geniculate nucleus of the cat visual system. *Neuroscience Letters, 361*, 83–85.

Podvigin, N., Jokeit, H., Pöppel, E., Chizh, A., & Kiselyeva, N. (1992). Stimulus dependent oscillatory activity in the lateral geniculate body of the cat. *Naturwissenschaften, 79*, 428–431.

Pöppel, E. (1970). Excitability cycles in central intermittency. *Psychologische Forschung, 34*, 1–9.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences, 1*, 56–61.

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B, 363*, 1887–1896.

Pöppel, E., & Logothetis, N. (1986). Neuronal oscillations in the human brain. *Naturwissenschaften, 73*, 267–268.

Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral system states: A hypothesis. *Naturwissenschaften, 77*, 89–91.

Radil, T., Mates, J., Ilmberger, J., & Pöppel, E. (1990). Stimulus anticipation in following rhythmic acoustical patterns by tapping. *Experientia, 46*, 762–763.

Sidiropoulos, K., Ackermann, H., Wannke, M., & Hertrich, I. (2010). Temporal processing capabilities in repetition conduction aphasia. *Brain and Cognition, 73*, 194–202.

Strasburger, H. (2001). Invariance of the psychometric function for character recognition across the visual field. *Perception & Psychophysics, 63*, 1356–1376.

Szelag, E., Szymaszek, A., Aksamit-Ramotowska, A., Fink, M., Ulbrich, P., Wittmann, M., et al. (2011). Temporal processing as a base for language universals: Cross-linguistic comparisons on sequencing abilities with some implications for language therapy. *Restorative Neurology and Neuroscience, 29*, 35–45.

Szelag, E., von Steinbüchel, N., & Pöppel, E. (1997). Temporal processing disorders in patients with Broca's aphasia. *Neuroscience Letters, 235*, 33–36.

Szymaszek, A., Sereda, M., Pöppel, E., & Szelag, E. (2009). Individual differences in the perception of temporal order: The effect of age and cognition. *Cognitive Neuropsychology, 26*, 135–147.

Szymaszek, A., Szelag, E., & Sliwowska, M. (2006). Auditory perception of temporal order in humans: The effect of age, gender, listener practice and stimulus presentation mode. *Neuroscience Letters, 403*, 190–194.

Tallal, P., Merzenich, M. M., Miller, S., & Jenkins, W. (1998). Language learning impairments: Integrating basic science, technology, and remediation. *Experimental Brain Research, 123*, 210–219.

Tallal, P., Miller, S., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*, 81–84.

Tewes, U. (1994). *HAWIE-R: Hamburg-Wechsler Intelligenztest für Erwachsene, Revision 1991*.

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*, 2993–3013.

Treutwein, B. (1995). Adaptive psychophysical procedures: A review. *Vision Research, 35*, 2503–2522.

Treutwein, B. (1997). YAAP: Yet another adaptive procedure. *Spatial Vision, 11*, 129–134.

Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2009). Perception of temporal order: The effects of age, sex, and cognitive factors. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition, 16*, 183–202.

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Wouters, J., & Ghesquière, P. (2011). Impairments in speech and nonspeech sound categorization in children with dyslexia are driven by temporal processing difficulties. *Research in Developmental Disabilities, 32*, 593–603.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*, 1293–1313.

Wittmann, M., & Fink, M. (2004). Time and language-critical remarks on diagnosis and training methods of temporal-order judgment. *Acta Neurobiologiae Experimentalis, 64*, 341–348.

Wittmann, M., & Szelag, E. (2003). Sex differences in perception of temporal order. *Perceptual and Motor Skills, 96*, 105–112.

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical band width in loudness summation. *The Journal of the Acoustical Society of America, 29*, 548–557.