Assessment of auditory temporal-order thresholds – A comparison of different measurement procedures and the influences of age and gender

Martina Fink\textsuperscript{a,}\textsuperscript{*}, Jan Churan\textsuperscript{a} and Marc Wittmann\textsuperscript{b}

\textsuperscript{a}Generation Research Program, Human Science Center, Ludwig-Maximilian University of Munich, Prof.-Max-Lange-Platz 11, 83646 Bad Tölz, Germany  
\textsuperscript{b}Department of Psychiatry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92037 – 0603 V, USA

Received 28 January 2005  
Revised 17 May 2005  
Accepted 23 May 2005

Abstract. Purpose: The relationship between auditory temporal-order perception and phoneme discrimination has been discussed for several years, based on findings, showing that patients with cerebral damage in the left hemisphere and aphasia, as well as children with specific language impairments, show deficits in temporal-processing and phoneme discrimination. Over the last years several temporal-order measurement procedures and training batteries have been developed. However, there exists no standard diagnostic tool for adults that could be applied to patients with aphasia. Therefore, our study aimed at identifying a feasible, reliable and efficient measurement procedure to test for auditory-temporal processing in healthy young and elderly adults, which in a further step can be applied to patients with aphasia. 

Methods: The tasks varied according to adaptive procedures (staircase vs. maximum-likelihood), stimuli (tones vs. clicks) and stimulation modes (binaural vs. alternating monaural) respectively. A phoneme-discrimination task was also employed to assess the relationship between temporal and language processing. 

Results: The results show that auditory temporal-order thresholds are stimulus dependent, age related, and influenced by gender. Furthermore, the cited relationship between temporal-order threshold and phoneme discrimination can only be confirmed for measurements with pairs of tones. 

Conclusion: Our results indicate, that different norms have to be established for different gender and age groups. Furthermore, temporal-order measurements with tones seem to be more suitable for clinical intervention studies than measurements with clicks, as they show higher re-test reliabilities, and only for measurements with tones an association with phoneme-discrimination abilities was found.

Keywords: Auditory-temporal processing, phoneme discrimination, aphasia, age difference, sex difference, diagnostics

1. Introduction

The association between temporal-processing mechanisms and language competence has been investigated over the last few decades. Studies have shown that pa-
tients with cerebral damage in the left hemisphere and aphasia [43, 59], children with specific language impairments [45], and children and adults with dyslexia [7] show deficits in identifying the correct order of two rapidly-presented, consecutive stimuli.

Based on findings showing that disturbed temporal-processing abilities are often correlated with phoneme identification or discrimination abilities, the association between temporal processing and language competence is discussed on the phonemic level. It is assumed that an analysis of temporal aspects of the speech signal is required to identify and discriminate certain phonemes [2, 8]. Temporal cues in speech sounds are formant transitions, characterized as short sound waveforms that change frequency across a time interval of ca. 40 ms. Consonant-vowel syllables, such as /ba/, /da/, and /ga/, vary according to their formant transitions, and it is hypothesized that this rapidly-changing signal has to be adequately analyzed to discriminate between these speech sounds. Another temporal element is the voice-onset time (VOT), the time between the burst and the onset of laryngeal pulsing that differentiates between voiced and unvoiced consonants, like /ba/ and /pa/ or /da/ and /ta/. Studies show that children with specific language impairments [47], children and adults with dyslexia [36], and patients with aphasia [38] have problems in discriminating stop-consonant vowel syllables. At the same time, increased temporal-order thresholds can be detected. This association supports the hypothesis that temporal-processing mechanisms are required for decoding spoken language [6, 39, 48].

The temporal order of two stimuli can be correctly identified by healthy subjects if their onset is separated by more than 20–40 ms [16, 25, 33]. The temporal-order threshold is defined as the minimum temporal interval between two auditory stimuli required for a subject to identify the correct order of two successive events. A model of a central temporal-processing mechanism located in the left-cortical hemisphere was proposed as the neurophysiological basis for identifying the temporal order of two separate stimuli [32, 33]. Several recent studies implicate age-related differences in temporal-processing mechanisms. Children below 10 years of age demonstrate higher temporal-order thresholds than adults [1, 19]. Additionally, it has to be mentioned that temporal-order thresholds in children show a wide range, between 30 ms and 200 ms. Studies reporting that older subjects have increased temporal-order thresholds support the assumption that temporal-processing mechanisms are age dependent [39]. This hypothesis is corroborated by reports implicating declines in various aspects of temporal processing in older listeners, such as identifying the temporal order of three-tone sequences [9], gap detection [37], and recognition performance of temporally-altered speech [11]. Another factor that was shown to influence temporal processing is gender. Several studies assessing different temporal-processing mechanisms showed sex differences in duration judgment [3, 34] and temporal-order detection [57].

Based on the hypothesis that temporal-processing mechanisms could form the basis of cognitive functions, such as language processing, several diagnostic methods have been developed over the last decade [26, 49]. Nonetheless, no standard diagnostic tool has been established to assess the auditory temporal-order threshold in adults that could also be used for clinical application assessing perceptual abilities of patients with aphasia [58]. Measurements to test the temporal-order threshold are usually performed by presenting two consecutive stimuli using different inter-stimulus intervals. Subjects have to judge the order of the presented stimuli, and the temporal-order threshold is determined by a psychophysical method. Instruments and procedures developed to assess the temporal-order threshold vary in different aspects, such as the psychophysical method, the stimuli employed, and the presentation mode.

1.1. Psychophysical method

One conventional procedure is the method of constant stimuli; several stimulus levels are pre-defined by the experimenter and presented repeatedly in randomized order to the participant. With this procedure, a wide range of stimulus levels can be covered. This allows psychophysical performance, as dependent on varying stimulus levels, to be completely characterized. However, if one is specifically interested in one single aspect of a psychophysical function, like the threshold, then the method of constant stimuli is rather inefficient because many stimuli are directed at regions other than those of interest [24]. Nevertheless, the method of constant stimuli is applied to assess temporal-order thresholds. In these studies, adequate temporal processing was variously defined, e.g., as 100% correct identifications of temporal order per stimulus level [6], or error rates were compared between short inter-stimulus intervals and long inter-stimulus intervals to judge temporal-order processing abilities [46].

Adaptive psychophysical procedures, in contrast, calculate the stimulus level (e.g. inter-stimulus inter-
val) of any trial based on the responses of the preceding trials. Therefore, adaptive procedures are suitable for measuring certain aspects of the psychophysical function, like the threshold, effectively and accurately [23]. Common adaptive measurements are staircase procedures, which decrease the stimulus level after a correct response and increase the stimulus level after an incorrect response. Thresholds are commonly estimated by averaging the reversal points obtained during the measurement. Simple up-down procedures target the 50% performance level of a psychometric function, whereas other performance levels can be obtained by transformed up-down procedures [24]. Using these procedures, stimulus levels are decreased following a defined number of consecutive correct response trials and are increased after each incorrect response. A study comparing different procedures suggests that three-step transformed up-down procedures, which track the 79.4% performance level, are more efficient than two-step transformed up-down procedures [21]. Different staircase procedures have been applied to test temporal-order processing [22,30].

In contrast to staircase procedures, where only the reversal points are averaged to estimate the threshold, maximum-likelihood procedures compute all trials to determine the threshold. While testing, stimulus levels in each trial are set according to the current best estimate of the threshold. After each trial, a psychometric function is generated based on the data points of previous trials using a maximum-likelihood fitting procedure [52]. Although in most studies using a fitting procedure only the threshold is estimated, other parameters of the psychometric function, such as the guessing rate, lapsing rate, and slope can also be estimated to provide a more precise description of subjects’ performance, and therefore increase the reliability and efficiency of measurements [23,54,56]. An adaptive measuring procedure based on maximum-likelihood parameter estimation, called YAAP (“Yet Another Adaptive Procedure”) [53], was implemented in a computer-aided program developed by Mates [26] to assess the temporal-order threshold. This tracking procedure estimates a threshold corresponding to 75% correct order discrimination. This method has been applied, for example, to children with and without specific language impairment and/or dyslexia [1,40] and to patients with aphasia [39].

1.2. Stimuli and stimulation modes

Stimuli used for assessing temporal-order processing can be divided into non-verbal stimuli and verbal stimuli (only non-verbal stimuli are discussed here – for a thorough discussion of non-verbal and verbal stimuli see [58]). In non-verbal stimuli, variations in physical properties, duration, and stimulation modes can be observed. In an alternating monaural stimulation mode, two click sounds are presented to the participants. One click is presented to the left ear and the other to the right ear. Subjects have to indicate the sequence of the clicks: left-right or right-left. Thus, this temporal-order task has also a spatial component. These click sounds usually have identical physical properties and are equal in length [15,18,25,29,39]. In a binaural condition, a sound is presented to both ears followed by a second, different sound with a certain inter-stimulus interval, and the subjects again have to indicate the order of the two stimuli. Studies employing binaural tasks commonly use pairs of tones, which differ in frequency and duration. For example, while Efron [6] used two sinusoidal tones with a fundamental frequency of 250 Hz and 2500 Hz and a duration of 10 ms, Talal and Piercy [46] used tones with a fundamental frequency of 100 Hz and 305 Hz and a duration of 75 ms. Note, that differences in the physical properties of the stimuli as well as different stimulation modes enable various cognitive strategies. Therefore, these parameters can also account for differences in temporal-order thresholds.

As these examples clearly show, there exist significant differences in stimulus presentation (binaural vs. alternating monaural), stimulus quality (clicks vs. tones), and duration of stimuli. Further clarification is necessary concerning the duration of stimuli. A stimulus duration of 1 ms, like in the alternating monaural condition, and an inter-stimulus interval of, for instance, 40 ms leads to a stimulus-onset asynchrony (SOA) of 41 ms. Using pairs of tones with 75-ms duration and the same inter-stimulus interval of 40 ms results in a SOA of 115 ms. Therefore, tones with a duration of 75 ms seem unsuitable for obtaining temporal-order thresholds in adults [58] because even with the lowest possible inter-stimulus interval of 1 ms, the SOA exceeds the time range of the temporal-order threshold discussed as being around 30 ms [31,33].

As mentioned above, new adaptive psychophysical methods have been developed to increase the efficiency and reliability of measurement. However, these procedures have not frequently been used to measure temporal-order processing. Especially in patients with aphasia or children with specific language impairments, it would be advantageous to use these methods. Due to a reduced number of trials, the exposure of the task
is shorter, and interfering factors, such as sustained attention should have less influence on subjects’ performance [58]. Moreover, reliability studies of measurements can not be found in the literature. A study which addressed the question of re-test reliability in children after one week and after four months achieved only low to moderate values with the alternating monaural stimulation method (1 week: $r = 0.53$; 4 months: $r = 0.46$) [1]. In clinical intervention studies where participants are assigned to groups based on their temporal-processing abilities, reliable measurements are necessary to avoid false assignments.

Based on studies showing a relationship between temporal-order processing and language processing on the phonemic level, the hypothesis of a temporal-processing deficit as an underlying cause for deficient language processing has been discussed for several years. As contradictory empirical results exist, the debate about a possible causal relationship is ongoing [42,60]. Unfortunately, no standard diagnostic tool has been established so far to test temporal-processing mechanisms in adult patients with aphasia. Moreover, no comparison between different auditory temporal-order measurements has been carried out to examine reliability and efficiency. Therefore, a feasible, reliable and efficient measurement procedure has to be identified first in healthy subjects, and then in patients with aphasia, to finally develop training methods in temporal-order judgment comparable to those developed for children with language-learning impairment [28,49].

The present investigation aimed at identifying a suitable procedure for obtaining the auditory temporal-order threshold in healthy subjects. Therefore, four different auditory temporal-order measurement procedures were compared by applying different stimuli (tones vs. clicks), respectively stimulation modes (binaural vs. alternating monaural), and psychophysical procedures (staircase vs. maximum likelihood). All measurements were repeated three times to gain information about the re-test reliability of these procedures. The influences of age and gender on temporal-order processing were also examined. Finally, a phoneme-discrimination task was employed to assess the relationship between temporal and language processing.

2. Materials and methods

2.1. Participants

In this study, 40 adults were assigned to two groups. Each group consisted of 20 participants (10 men and 10 women). The first group included elderly subjects, aged 55–70 (mean age 61.7), and the second group consisted of younger subjects, aged 20–35 (mean age 25). All participants were tested for normal hearing and had no history of neurological disease. To exclude subjects with hearing deficits, hearing function for both ears was assessed using pure-tone audiometry (Audiometer MA 15, Maico Diagnostic GmbH). The adaptive procedure used frequencies ranging from 500 Hz to 4000 Hz (500, 750, 1000, 1500, 2000, 3000, 4000 Hz) and a dB range from $-10$ dB to $100$ dB in steps of 5 dB. A hearing level of 10 dB, for example, means that the subjects requires 10 dB more than the average for tone detection. The criteria for admission to the study were a hearing level below 30 dB for all frequencies tested and differences in hearing level between the two ears below 20 dB. Due to this screening procedure 5 subjects were not included in the study.

Three test sessions on three different days were conducted with all participants. Temporal-order threshold measurements, as well as phoneme discrimination, was repeated every session. All subjects were tested individually; a single auditory temporal-order measurement lasted for approximately 10 minutes, a test session on one day lasted approximately one hour. Subjects were paid for their participation in the study.

2.2. Temporal-order threshold

The auditory temporal-order threshold is defined as the minimum temporal interval between two auditory stimuli that must exist before a person is able to identify the correct order of two successive events. In the present investigation, the threshold corresponds to 75% correct order discrimination. Auditory temporal-order thresholds were assessed using computer-aided systems (Pentium 3 processor; soundblaster audigy soundcard). Four measurements were taken using different stimuli, adaptive procedures, and stimulation modes. Table 1 displays the four measurements.

2.3. Stimuli

All stimuli used for the temporal-order measurements were pairs of acoustic events. The stimuli were generated with the program Cool Edit 2000 (sampling rate 44100 Hz, 16-bit). In the present investigation, two stimulus conditions were applied. Two of the four measurements used clicks, and the other two measurements used sinusoidal tones.
Table 1

<table>
<thead>
<tr>
<th>Test name</th>
<th>Stimulus</th>
<th>Adaptive rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>staircase</td>
<td>clicks</td>
<td>Transformed up-down method [24]</td>
</tr>
<tr>
<td>YAAP</td>
<td>clicks</td>
<td>YAAP [53]</td>
</tr>
<tr>
<td>staircase</td>
<td>tones</td>
<td>Transformed up-down method [24]</td>
</tr>
<tr>
<td>YAAP</td>
<td>tones</td>
<td>YAAP [53]</td>
</tr>
</tbody>
</table>

Clicks were rectangular pulses of 1 ms duration, presented in an alternating monaural stimulation mode (one click to each ear). Tone stimuli consisted of a low tone (800 Hz) and a high tone (1200 Hz). The duration was 10 ms with 1 ms rise-and-fall time, and the tones were presented in a binaural stimulation mode (a tone is presented to both ears followed by a second tone to both ears). The inter-stimulus interval varied during the experiments according to a certain adaptive rule (see below).

For all measurements, each listening trial included three warning signals prior to the stimulation to focus attention on the task. The time between warning signal and stimulus was 1500 ms. All stimuli were presented with 82.3 dB SPL via headphones (SONY MDR-CD 480). The order of stimulus presentation was randomized.

2.4. Psychophysical methods

For the measurement of the auditory temporal-order threshold, a pair of stimuli was presented to the participant, and the subject had to indicate the order of the two acoustic stimuli (clicks or tones) in a two-alternative, forced-choice task. At the beginning of a test session, practice trials were carried out to demonstrate the task and make sure the subjects understood the instructions. In this training phase, trials with an inter-stimulus interval of 300 ms were repeated several times to ensure that subjects could discriminate the two stimuli. Thresholds were obtained using two different adaptive procedures for changing the inter-stimulus interval on the basis of the subject’s responses.

2.5. Transformed up-down procedure

A 3-step, transformed up-down procedure [24] was used, which decreases the inter-stimulus interval following three consecutive correct response trials and increases the inter-stimulus interval after each incorrect response. One test session included 40 trials with an initial inter-stimulus interval of 100 ms and an initial step size of 30 ms. During measurement, the step size was halved at each reversal (increase vs. decrease of inter-stimulus intervals), and the minimum step size was defined as 5 ms. Inter-stimulus intervals for click stimuli ranged from 10 ms to 200 ms. For tone stimuli, the range was 1 ms to 200 ms. The threshold was defined as the average of the last five reversals. This tracking procedure estimates a threshold corresponding to 79.4% correct order discrimination.

2.6. YAAP

Alternatively, a maximum-likelihood based algorithm – YAAP [53] – was used. In this test, the stimulus in each trial is set at the current best estimate of the threshold. This tracking procedure estimates based on a logistic psychometric function a threshold corresponding to 75% correct order discrimination. YAAP test sessions included two test 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 (125 ms) and proceeded in equal steps of 20% to a specified lower limit (clicks: 10 ms; tones: 1 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 [52]. Parameters of the underlying logistic psychometric function were specified as follows: guessing rate: 50%; lapsing rate: 1%; spread: 4.5.

The number of trials in the YAAP algorithm was 39.14 in average. Therefore, the number of trials is comparable in both algorithms used.

1 The implementation of the algorithm used in this study is based on the implementation of Mates et al. [26] and was adapted under supervision of Hans Strasburger and Bernhard Treutwein.

2 Initial estimate of a parameter describing the spread or inverse steepness (1/slope) of the function [26].
2.7. Phoneme discrimination

In a phoneme-discrimination task, subjects had to discriminate between the words /danken/ (to thank) and /tanken/ (to fuel) ([20]; for a more accurate description of the stimulus generation). The task included ten stimuli varying in voice-onset time of the initial stop consonant. Stimuli were created by manipulating the voice-onset time of the stop consonant in a naturally-spoken /tanken/. This resulted in phonemes with voice-onset times ranging between 0 ms and 90 ms in 10 ms steps. During a test session, every stimulus was presented ten times in randomized order using the method of constant stimuli. The words were presented via headphone with 82.3 dB SPL. Participants responded by pointing to a respond card (see Fig. 1) and the responses were documented by the experimenter.

2.8. Statistical analysis

As the psychophysical procedures used in this study track different performance levels (Transformed up-down procedure: 79.4%; Y AAP: 75%), a best-fit procedure was implemented to make the measurements comparable. Psychometric functions were fitted using psignifit version 2.5.41 (see http://bootstrap-software.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann and Hill [56]. This procedure uses a logistic psychometric function model and estimates three parameters: (1) the threshold, corresponding to 75% correct order discrimination, (2) the slope, and (3) the lapsing rate [41]. This procedure was also obtained for analyzing phoneme-discrimination abilities. Thresholds corresponding to 50% identification of the word /tanken/ were defined as the category boundary.

As the obtained temporal-order thresholds indicate the inter-stimulus interval corresponding to 75% correct order discriminations, SOAs were assigned by adding the stimulus duration to the thresholds (clicks: 1 ms; tones: 10 ms). An inter-stimulus interval of 20 ms e.g. represents a SOA of 21 ms for the measurements with clicks and 30 ms for the measurements with tones. As several subjects showed nearly 100% correct temporal-order identification at the lowest inter-stimulus interval in the tone condition (a SOA of 11 ms), thresholds were estimated for these subjects at half of the smallest used SOA. Threshold values of 5.5 ms were assigned to 43% of temporal-order measurements in younger subjects and 13% of temporal-order measurements in older subjects.

The Kolmogorov-Smirnov test was applied to the data to test for normal distribution. Results showed that 58.3% of the data were not normally distributed. Therefore non-parametric methods were used to analyze the data. To compare the thresholds from the three test sessions, Friedman and Wilcoxon tests were applied. For comparisons of different stimuli, and the different psychophysical procedures the Wilcoxon test was used. To analyze group differences, Mann-Whitney U-Tests were implemented. Spearman rank correlations were used to test associations among different measurements. Bonferroni corrected alpha levels were applied for repeated comparisons.

3. Results

3.1. Differences among the three test sessions

Results of the temporal-order thresholds obtained with the four different procedures are displayed in Fig. 2. The figure shows the median of temporal-order thresholds obtained from the four different procedures over all participants. As can be seen in the figure, thresholds obtained with tones (triangles) were lower for all test sessions in both algorithm conditions (detailed description see below).

A comparison of the thresholds obtained in the three test sessions revealed different results for the four procedures. Staircase_click and YAAP_click showed no significant differences among the three test sessions. Although the analysis of YAAP_click showed a significant difference among test sessions, the following pairwise comparison with the Wilcoxon test yielded no significant differences among the temporal-order thresholds of the three test sessions (Friedman test: staircase_click: $df = 2$, $\chi^2 = 1.54$, $p = 0.462$; YAAP_click: $df = 2$, $\chi^2 = 6.40$, $p = 0.041$ – Wilcoxon test: YAAP_click: 1–2: $z = -0.57$, $p = \ldots$).
Table 2

<table>
<thead>
<tr>
<th></th>
<th>Test Sessions 1&amp;2</th>
<th>Test Sessions 1&amp;3</th>
<th>Test Sessions 2&amp;3</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>staircases_click</td>
<td>0.724**</td>
<td>0.523**</td>
</tr>
<tr>
<td></td>
<td>YAAP_click</td>
<td>0.614**</td>
<td>0.436**</td>
</tr>
<tr>
<td></td>
<td>staircases_tone</td>
<td>0.794**</td>
<td>0.834**</td>
</tr>
<tr>
<td></td>
<td>YAAP_tone</td>
<td>0.781**</td>
<td>0.768**</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2. Comparison of different stimulus types

The comparison of thresholds obtained with the staircase algorithm and with different types of stimuli revealed significant differences between the thresholds for test sessions 2 and 3 after Bonferroni correction (test session 1: $z = -1.95, p = 0.051$; test session 2: $z = -2.47, p = 0.013$; test session 3: $z = -2.90, p = 0.004$). Data analysis of thresholds obtained with the YAAP algorithm with clicks and with tones, respectively, showed only for test session 3 significant differences between the temporal-order thresholds after Bonferroni correction (test session 1: $z = -1.61, p = 0.106$; test session 2: $z = -2.04, p = 0.041$; test session 3: $z = -3.08, p = 0.002$). Figure 2 shows the median, 25 percentile, and 75 percentile of temporal-order thresholds for the four procedures, arranged according to psychophysical procedures.
3.3. Comparison of different psychophysical methods

The comparison of temporal-order thresholds obtained with clicks using the two different algorithms showed no significant differences over the three test sessions (test session 1: \( z = -0.57, p = 0.572 \); test session 2: \( z = -1.05, p = 0.293 \); test session 3: \( z = -0.89, p = 0.372 \)). Comparison of thresholds obtained with pairs of tones using the two types of algorithms yielded no significant differences, either (test session 1: \( z = -0.32, p = 0.751 \); test session 2: \( z = -0.57, p = 0.567 \); test session 3: \( z = -0.86, p = 0.389 \)).

Rank correlations between threshold values obtained with different psychophysical procedures but with the same stimuli revealed significant high correlation coefficients, as can be seen in Table 3.

3.4. Age-related differences

Two different age groups were tested in this study. One group of younger subjects, aged 20–35, and another group of older subjects, aged 55–70. Group comparisons of temporal-order thresholds were analyzed for the four measurement procedures separately. Higher temporal-order thresholds were detected in older subjects in all measurements. Although thresholds of older subjects obtained with clicks were higher than the thresholds of younger subjects, only with the staircase algorithm the third test session revealed significant differences after Bonferroni correction (staircase_click: test session 1: \( z = -1.49, p = 0.141 \); test session 2: \( z = -1.34, p = 0.189 \); test session 3: \( z = -2.68, p = 0.007 \); YAAP_click: test session 1: \( z = -1.25, p = 0.221 \); test session 2: \( z = -1.10, p = 0.283 \); test session 3: \( z = -2.08, p = 0.038 \)).

In contrast, the group comparison of temporal-order thresholds obtained with tone stimuli yielded different results. The analysis of staircase_tone showed significant differences between the thresholds of the two age groups in all three test sessions after Bonferroni correction (test session 1: \( z = -3.16, p = 0.001 \); test session 2: \( z = -3.88, p < 0.001 \); test session 3: \( z = -3.21, p = 0.001 \)). The two age groups showed also significant differences in all three test sessions of YAAP_tone (test session 1: \( z = -3.63, p < 0.001 \); test session 2: \( z = -2.90, p = 0.004 \); test session 3: \( z = -3.63, p < 0.001 \)). Figure 3 shows the temporal-order thresholds for the two age groups arranged according to the four different measurement procedures.

It should be noted that several subjects had nearly 100% correct temporal-order identification at the lowest inter-stimulus interval in the tone condition. As has been described previously, thresholds for these subjects were estimated. Threshold values of 5.5 ms were assigned to 43% of temporal-order measurements in younger subjects and 13% of temporal-order measurements in older subjects.

3.5. Sex-specific differences

Twenty female and twenty male subjects were included in the tests. Group comparisons of temporal-order thresholds were analyzed for the four measurement procedures separately. Results for staircase_click showed significant differences in test session 1, whereas in test sessions 2 and 3, no significant differences between the temporal-order thresholds of male and female listeners were observed after Bonferroni correction (test session 1: \( z = -2.93, p = 0.003 \); test session 2: \( z = -2.16, p = 0.031 \); test session 3: \( z = -1.03, p = 0.314 \)). Statistical analysis of temporal-order thresholds obtained with YAAP_click showed significant differences between male and female listeners only in test session 1. In test sessions 2 and 3, no significant differences were recorded after Bonferroni correction (session 1: \( z = -2.67, p = 0.007 \); test session 2: \( z = -1.83, p = 0.070 \); test session 3: \( z = -0.73, p = 0.478 \)). Comparing thresholds obtained with staircase_tone and YAAP_tone, the results revealed no significant differences between male and female listeners (staircase_tone: session 1: \( z = -0.65, p = 0.531 \); session 2: \( z = -0.47, p = 0.644 \); session 3: \( z = -0.65, p = 0.537 \); YAAP_tone: ses-

<table>
<thead>
<tr>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>staircase_click</td>
<td>0.838**</td>
<td>0.745**</td>
</tr>
<tr>
<td>YAAP_click</td>
<td>0.860**</td>
<td>0.873**</td>
</tr>
</tbody>
</table>

Rank correlation coefficients between temporal-order thresholds obtained with different psychophysical procedures but the same stimuli; Conventions as in Table 2.
Fig. 3. Median, 25 percentile and 75 percentile of temporal-order thresholds in ms for the two age groups obtained with four different measurement procedures. Upper left: staircase algorithm and clicks, Lower left: staircase algorithm and tones, Upper right: YAAP algorithm and clicks, Lower right: YAAP algorithm and tones; ◦/○: younger subjects, ▽/▼: older subjects; Conventions as in Fig. 2.

3.6. Association between temporal processing and phoneme discrimination

Based on the assumption that an analysis of temporal elements of the speech signal, like the VOT, is necessary to correctly identify and discriminate stop-consonants, higher temporal-order thresholds are expected to be associated with a delayed category boundary (= at longer VOTs) between voiced and voiceless phonemes. Therefore, rank correlations between temporal-order thresholds and phoneme discrimination were carried out to test for an association between these two processing mechanisms. The category boundaries of the three measurements for both age groups are displayed in Fig. 5. Older subjects showed significantly higher values than the younger subjects in all three measurements after Bonferroni correction (measurement 1: \( z = -2.42, p = 0.014 \); measurement 2: \( z = -3.64, p < 0.001 \); measurement 3: \( z = -3.84, p < 0.001 \)). In contrast, no effects of gender can be seen (measurement 1: \( z = -1.63, p = 0.105 \); measurement 2: \( z = -1.31, p = 0.192 \); measurement 3: \( z = -1.10, p = 0.277 \)).

As the comparison between category boundaries (between /danken/ and /tanken/) of the three test sessions showed no significant difference (\( p = 0.705 \)), and correlations showed values at a significance level of \( p < 0.01 \) (sessions 1–2: \( r = 0.723^{**} \); sessions 1–3: \( r = 0.683^{**} \); sessions 2–3: \( r = 0.910^{**} \)), the values of the different sessions were averaged. Results showed
Fig. 4. Median, 25 percentile and 75 percentile of temporal-order thresholds in ms for the two gender groups obtained with four different measurement procedures. Upper left: staircase algorithm and clicks, Lower left: staircase algorithm and tones, Upper right: YAAP algorithm and clicks, Lower right: YAAP algorithm and tones; ○ ●: men, ▽ ▼: women; Conventions as in Fig. 2.

no significant correlations between averaged thresholds obtained with click sounds and phoneme discrimination. In contrast, significant correlations between phoneme discrimination and averaged temporal-order thresholds were obtained with pairs of tones. Correlation coefficients are shown in Table 4.

4. Discussion

This study was designed to compare different auditory temporal-order measurements and obtain information about their re-test reliability. Effects of age and gender on temporal-order thresholds were also examined. Finally, the association between temporal processing and phoneme discrimination was assessed.

4.1. Re-test reliability

The comparison of the three test sessions reveals different results for the two different types of stimuli. Comparison of temporal-order thresholds obtained with clicks show no significant differences among the three test sessions. However, correlation coefficients show moderate values, indicating intra-individual variance over the three test sessions. As only one single measurement is commonly used to determine temporal-order thresholds, this result is relevant for clinical-intervention studies. In such studies, where participants are assigned to groups based on their temporal-processing abilities, group comparisons can be biased by false assignments. Therefore, it seems suitable to increase the number of temporal-order measurements with clicks and average the results to improve the va-
lidity. However, two problems have to be considered concerning this recommendation. First, increasing the number of temporal-order measurements is not easy in clinical settings due to time constraints. Second, sensitivity and specificity is influenced by an increase in the number of measurements. Multiple measurements decrease test specificity, but on the other hand increase test sensitivity. Therefore, when using click stimuli in a diagnostic setting, we propose the conduction of two to three measurements. The average of the three measurements then should be evaluated to reduce false positive assignments to a pathologic group.

The comparison of the three consecutively assessed temporal-order thresholds obtained with pairs of tones shows significant differences among the measurements for the staircase algorithm, as well as for YAAP. The values decrease from session to session. Looking at the correlation coefficients, high values can be found, indicating stable relationships between participants. This latter finding is more important for clinical assessment, since the number of false group assignments of participants based on their temporal-processing abilities in one measurement would be lower for tones than for clicks. Nevertheless, the changes in absolute values of the tone thresholds over the three sessions must be considered e.g. by separate norms for different sessions.

As discussed earlier, reliability studies can hardly be found in literature. The reliability of measurement procedures is, however, an important factor, especially in neuropsychological studies. The results of this study emphasize the need for reliability studies of existing test batteries for the measurement of auditory temporal-order perception.

4.2. Learning

In recent years not only diagnostic tools for the measurement of temporal processing, but also training procedures have been developed. Based on the assumption that an improvement of temporal-processing abilities will lead to an improvement in phoneme discrimination, children with specific-language impairments or patients with aphasia were trained in their temporal-order threshold [28,38]. As these studies show, temporal-order thresholds can be improved by a computer aided feedback training. This finding points out the clinical relevance of this research topic. In the present investigation healthy adults showed an improvement in their temporal-order thresholds obtained with pairs of tones over the three test sessions even without feedback. However, no changes in the phoneme-discrimination task were observed over the three test sessions. Therefore, it has to be noted that the improvement in the temporal-order task may not only be attributed to a training of temporal processing. The improvement of cognitive strategies or attentional adaptation to the task has to be considered as well. In this case, no changes in the phoneme-discrimination task would be expected.

4.3. Comparison of different stimuli

The results show that there are significant differences in temporal-order thresholds obtained with different kinds of stimuli. Thresholds obtained with pairs of tones are significantly lower than thresholds obtained with clicks. For test sessions 2 and 3, significant differences can be noted for the staircase procedures. For the YAAP algorithm session 3 reveals significant differences. In addition, no clear association between temporal-order thresholds obtained with the two different stimuli can be identified by the correlation coefficients. These results indicate stimulus-dependent processing of temporal order.

Previous studies indicate that different problem-solving strategies can influence temporal-processing tasks [13,44]. Moreover, it has been discussed that two different processing strategies can be applied in the perception of sound sequences. In an analytic processing mode, first the separate components are identified, and afterwards the temporal order is established. In a holistic processing mode, the temporal order is assigned by global pattern recognition [55,57]. Interestingly, in the second processing mode, there is no need to identify the individual tones to determine the temporal order. Subjects who were trained to use this holistic processing mode showed better performance [4]. In the present study, presentation of two tones with different fundamental frequencies facilitated the use of global-pattern recognition. As the two possible combinations of sounds generate two different spectral patterns (high-to-low; low-to-high), the order of the two sounds can be identified by analyzing the perceived tonal sequence. Concordant with the cited results, participants in the present investigation showed lower temporal-order thresholds with pairs of tones than with clicks.

In addition, results from single-neuron-activity recordings in animals provide interesting indications for processing frequency-modulated sounds (low-to-high and high-to-low frequencies). Results indicate that there exist neurons in the posterior auditory cortex
Table 4
Rank correlation coefficients between the category boundary (phoneme discrimination) and temporal-order thresholds; Conventions as in Table 2

<table>
<thead>
<tr>
<th></th>
<th>staircase_click</th>
<th>YAAP_click</th>
<th>staircase_tone</th>
<th>YAAP_tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneme discrimination</td>
<td>0.106</td>
<td>0.021</td>
<td>0.423**</td>
<td>0.405**</td>
</tr>
</tbody>
</table>

Fig. 5. Median, 25 percentile and 75 percentile of category boundaries in ms for the two age groups. ◦/●: younger subjects, ▽/▼: older subjects; Conventions as in Fig. 2.

which are selective for the direction of frequency-modulated sounds [27,35,50,51]. Assuming neurons selective for the direction of frequency modulations provides a possible explanation for the dissociation of temporal-order thresholds obtained with different kinds of stimuli in the present investigation. If the temporal order of two sounds with different frequencies is processed by selective neurons that process frequency-modulated sounds, the proposed central temporal-processing mechanism located in the left cortical hemisphere would not serve as a neurophysiological basis for identifying the temporal order of these specific stimuli. However, this central temporal-processing mechanism can be assumed to be the neurophysiological basis for identifying the temporal order of two clicks. Therefore, it is hypothesized that two different kinds of temporal-processing mechanisms are activated by the two types of stimuli.

4.4. Comparison of different psychophysical procedures

As the two psychophysical procedures applied in this investigation track different threshold levels (staircase: 79.4%; YAAP: 75%), a best-fit procedure was implemented to make the measurements comparable. The adaptive procedures were used here to collect the data near the individual threshold. Results reveal no significant differences in temporal-order thresholds obtained with different psychophysical procedures. Additionally, high correlation coefficients indicate an association between these measurements. These results suggest that the transformed up-down procedure and the YAAP algorithm are both suitable to obtain comparable temporal-order thresholds.

Nevertheless, one has to keep in mind that a best-fit procedure was applied in the present investigation to make a comparison possible. As mentioned above, different methods define the threshold at different performance levels. This has to be considered when comparing absolute threshold values across different psychophysical procedures.

4.5. Age-related differences

In the present investigation, the group of older listeners revealed higher temporal-order thresholds in all recorded measurements. For measurements obtained with clicks, the mean difference between younger and older subjects was 14.29 ms. Statistical analysis revealed significant group differences with the clicks only for the third measurement with the staircase procedure.

With the same subjects, a more prominent effect of age for temporal-order measurements with pairs of tones emerged. In this condition, the mean difference
between the two groups was 48.45 ms. In each of the three measurements significant group differences were found.

As stated previously, an age-related decline in temporal processing has been reported in several studies. The results of the present investigation corroborate these findings. Although the effects of age have often been documented, the cause of this decrease in performance is still unclear. Several authors propose that age-related declines in temporal processing depend on the processing speed of the brain [9,11,37]. Studies showing that older listeners exhibit deficits in the processing of dynamic acoustic cues of the speech signal, performing with poorer speech-sound identification [5,10], provide an interesting indication of why the older participants in our study have greater difficulties in temporal-order detection with tones. As the pairs of tones are characterized by a variation in frequency over time, they are similar to the dynamic acoustic cues of speech sounds (e.g. formant transitions). Decreased temporal-order processing of two tones could, therefore, occur due to a decline in the processing of dynamic acoustic structures. This hypothesis is supported by the result, that the group of older subjects had a category boundary at a significantly longer VOT than the younger subjects, indicating that they show a different processing of dynamic acoustic structures as compared to the younger participants.

### 4.6. Sex-specific differences

The results of the present investigation showed significantly lower temporal-order thresholds obtained with clicks in men than in women. The mean difference between men and women was 14.01 ms. For both psychophysical algorithms the first sessions revealed significant group differences. These results corroborate previous findings indicating lower thresholds in men in an alternating monaural stimulation mode [57]. Findings concerning temporal-order thresholds obtained with tones showed different results. No group comparisons between men and women reached statistical significance with tones.

Sex-specific differences in temporal-order thresholds obtained with clicks can be discussed in terms of neuro-anatomical differences. Reports indicate a stronger asymmetry in the cortical organization of the brains of men and a clearer lateralization of cognitive functions [12,14,17]. It has been discussed that a more accurate spatio-temporal representation of the two click sounds due to a clearer lateralization improves the temporal left-right discrimination [57]. As the alternating monaural temporal-order task has also a spatial component (one click to each ear), sex-specific differences are not only attributable to temporal aspects of the task.

In the second procedure in this study, a binaural stimulation paradigm was applied. The results showed no significant differences between men and women in this condition. Results of the present investigation showed that auditory temporal-order thresholds are stimulus dependent. We hypothesized that two different kinds of temporal-processing mechanisms are activated by the two different stimuli and that the temporal order of two sounds with different frequencies is probably processed by neurons selective for frequency-modulated sounds [27,35,50,51]. Results showing no significant differences between men and women in the binaural stimulation mode corroborate this hypothesis. If the temporal order of two tones is processed by selective groups of neurons, neuro-anatomical differences in the symmetry of cortical organization between men and women should not influence the temporal-order processing of two tones.

### 4.7. Association between temporal processing and phoneme discrimination

As described above, the association between temporal processing and language competence is discussed on the phonemic level. The analysis of temporal aspects of the speech signal is hypothesized as a basis for identifying and discriminating certain phonemes. Based on this assumption and findings that older persons show increased temporal-order thresholds [39] and reduced speech-sound identification [5,10], an association between the temporal-order threshold and the category boundary was expected. Results showed no significant correlations between thresholds obtained with clicks and phoneme-discrimination abilities. In contrast, significant positive correlations between thresholds and the category boundary were recorded for the binaural condition with pairs of tones. One possible explanation lies in the different physical properties of the stimuli. The pairs of tones applied in this study generated frequency transitions and were, therefore, more similar to the physical properties of speech sounds. As mentioned before, the debate about a possible causal relationship between temporal processing and phoneme discrimination is ongoing. Studdert-Kennedy and Mody [42] assume that the identification of spectral and not temporal components is necessary for phoneme identification. In addition, Wright et al. [60] found that chil-
dren with specific language impairments showed not a general deficit in the perception of rapidly presented sounds. Instead, their perceptual deficits depended on the temporal and spectral sound contexts.

Our results show that the detection of the temporal order of spectral elements is associated with phoneme-discrimination abilities. The detection of the temporal order of two clicks that are presented in an alternating monaural stimulation mode, however, is not correlated with phoneme discrimination. To sum up, our study supports the hypothesis of a relationship between a psychophysical threshold measure of non-verbal auditory temporal-order detection and categorical perception of speech stimuli based on voice-onset time differences.

5. Summary

A main finding of the present investigation is that temporal-order thresholds clearly depend on the physical properties of the stimulus. Different processing for the identification of temporal order could be one possible explanation for the significant differences between thresholds obtained with clicks and tones. Results showing that there are neurons that are specific for frequency modulation corroborate the hypothesis that two different kinds of temporal-processing mechanisms are required for the two stimuli types. Our results show that auditory temporal-order thresholds are stimulus dependent, age related, and influenced by gender. Therefore, different norms have to be established for different age and gender groups.

Additionally, re-test reliabilities indicate that an increase in the number of temporal-order measurements would be helpful to improve the validity, especially of measurements with clicks. Finally, the assumed association between temporal-order thresholds and phoneme discrimination can only be confirmed for measurements with tones.

Our results indicate that temporal-order measurements with tones are more suitable for clinical intervention studies than measurements with clicks according to several reasons. First, their re-test reliability is higher than that of measurements with clicks. This is important in clinical settings, where multiple measurements are difficult to conduct due to time constraints. Second, a significant positive correlation between temporal-order measurements and phoneme-discrimination abilities could only be confirmed for measurements with pairs of tones. According to this investigation, new questions have to be addressed in clinical-intervention studies:

1) Is stimulus-dependent processing of auditory-temporal order also existent in patients with aphasia? Different training procedures would then have to be developed and tested for efficiency.

2) Is the stronger association between temporal-order thresholds obtained with tones and phoneme-discrimination abilities also existent in patients with aphasia?

The present investigation provided a first step in the identification of a feasible, reliable and efficient measurement procedure to test for auditory-temporal processing in adults and patients with aphasia. However, more work has to be done and the next step will be to compare different auditory-temporal order measurements in patients with aphasia and to relate them to phoneme-discrimination abilities.

Acknowledgements

The authors wish to thank Sebastian Rummel for validating the physical properties of the non-verbal stimuli electronically, and Hans Strasburger for his help in supervising the implementation of the psychophysical algorithm in our temporal-order threshold program. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF), project number FKZ 01GZ0301.

References


M. Kiss, Der Zusammenhang von Zeitverarbeitung und Sprache bei Patienten mit flüssigen Aphasien [The association of temporal processing and language processing in patients with fluent aphasia], Dr. Hut Verlag, München, 2002.


P. Tallal, S.L. Miller, G. Bedi, G. Byma, X. Wang, S.S. Nagarajan, C. Schreiner, W.M. Jenkins and M.M. Merzenich, Language Comprehension in language-learning impaired chil-


