The development of a fast method for recording Schroeder-phase masking functions

Sarah Rahmat[1], Greg A. O'Beirne[2]

1. *Department of Audiology & Speech Language Pathology, International Islamic University Malaysia*
   Email: sarahrahmat@iium.edu.my   Phone: +6095716400 ext 3305
2. *Department of Communication Disorders, University of Canterbury, New Zealand*

Abstract

Schroeder-phase masking complexes have been used in many psychophysical experiments to examine the phase curvature of cochlear filtering at characteristic frequencies, and other aspects of cochlear nonlinearity. In a normal nonlinear cochlea, changing the "scalar factor" of the Schroeder-phase masker from -1 through 0 to +1 results in a marked difference in the measured masked thresholds, whereas this difference is reduced in ears with damaged outer hair cells. Despite the valuable information it may give, one disadvantage of the Schroeder-phase masking procedure is the length of the test – using the conventional three-alternative forced-choice technique to measure a masking function takes around 45 minutes for one combination of probe frequency and intensity. As an alternative, we have developed a fast method of recording these functions which uses a Békésy tracking procedure. Testing at 500 Hz in normal hearing participants, we demonstrate that our fast method: i) shows good agreement with the conventional method; ii) shows high test-retest reliability; and iii) shortens the testing time to 8 minutes.

Key words: psychoacoustic methods; Békésy tracking; Schroeder-phase masking; cochlear nonlinearity

1. Introduction

The classical model of masking known as the power spectrum model suggests that the amount of masking is determined by the amount of noise passing through the auditory filter that is centred close to the frequency of the signal (Patterson and Moore, 1986). Increasing noise bandwidth will increase the amount of noise passing through the auditory filter as long as the noise bandwidth does not exceed the critical bandwidth. According to the power spectrum model, the amount of masking is determined by the frequency and intensity of the masker, and the phase of the input stimulus has been considered to be of little perceptual significance. This, however, is contradicted by findings in experiments over the past 20 years that use Schroeder-phase harmonic complexes as the masker.

Schroeder-phase harmonic complexes (Schroeder, 1970) have been used in many psychophysical experiments to examine the phase curvature of the auditory filter at characteristic frequencies, as well as other aspects of cochlear nonlinearity (Gifford et al., 2008; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Recio & Rhode, 2000; Summers & Leek, 1998). These complexes, which have identical amplitude spectra but which differ in their phase spectra (see Figure 1), have been observed to stimulate different patterns of basilar membrane excitation (Recio & Rhode, 2000). They also have different masking abilities when presented together with tones in normal healthy cochleae, producing different masked thresholds (Gifford et al., 2008; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Summers & Leek, 1998).

A formula for the construction of Schroeder-phase harmonic complexes is shown in Equation 1 (Carlyon & Datta, 1997; Kohlrausch & Sander, 1995; Recio & Rhode, 2000, Summers & Leek, 1998):

$$m(t) = \sum_{n=n1}^{n2} A_0 \sin(2\pi n f_0 t + \theta_n)$$

(Equation 1)

Different phases (θ_n) may be generated by changing the scalar factor (C) of the masker according to the alteration of Schroeder's (1970) original formula by Lentz and Leek (2001):

$$\theta_n = C\pi n(n+1)/N \, , -1 \leq C \leq 1$$

(Equation 2)

The difference in masking effectiveness at different scalar factors of Schroeder harmonic complexes is known as the 'phase effect'. The phase effect is quantified by the difference between maximum and minimum masked thresholds across different scalar factors and will be used as one of the important measures of Schroeder-phase masking in this study. The phase effect has been observed to decrease in participants with sensorineural hearing loss and at very low and high intensity sound levels (Gifford et al., 2008; Recio & Rhode, 2000; Summers & Leek, 1998).

There are several theories that may explain the underlying mechanisms of the phase effect. It is believed that the phase effect is due to different modulation patterns produced at the output of the auditory filter, due to interaction between the phase curvature of the Schroeder harmonic complexes and the negative phase curvature of the basilar membrane (Kohlrausch & Sander, 1995). That is, when Schroeder-phase complexes have the same magnitude but an opposite phase curvature from that of the basilar membrane, the phase differences between the adjacent components will be flattened out by the auditory filter to produce a peaky output in the time domain (Kohlrausch & Sander, 1995; Lentz & Leek, 2001). This peaky output of the internal waveform contains a region of low masker amplitude, enabling the signal to become audible (refer to Figure 1 of Summer & Leek, 1998). This phenomenon is known as 'listening in the gap', which is more pronounced when Schroeder-phase complexes with positive scalar factors are used as the masker (e.g. C = +1), resulting in less effective masking, and producing a low masked threshold (Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2004). Unlike Schroeder-phase complexes with positive scalar factors, those with negative scalar factors (e.g. C = -1) produce a more uniform output over time, which makes them more effective maskers and results in higher masked thresholds. This different pattern of interaction between the basilar membrane and the different phases of the Schroeder-phase masker is believed to contribute to the phase effect mechanism.

In addition, involvement of cochlear nonlinearity via ‘phasic suppression’ is also believed to contribute to the different masking effectiveness of different phases of Schroeder-phase complexes (Summers, 2000). This phasic suppression contributes more when using positive-phase Schroeder harmonic complexes than with negative-phase ones (Summers, 2000). That is, the active process / nonlinear cochlear processing suppresses the gain towards tones during the high amplitude regions of the internal response, and boosts the frequency-selective gain towards the tone during the low amplitude regions of the internal response of the masker. This phasic suppression benefits tone detection in the presence of Schroeder-phase maskers with positive scalar factors as the internal response contains periods/regions of low amplitude (valleys), but not in the presence of those with negative scalar factors as the internal response is more flat. This results in large masking differences between the two conditions, and therefore a large phase effect. The involvement of active mechanisms in explaining the phase effect was supported by findings from several studies: It has been observed that phase effect was level dependent, being reduced at both low and very high presentation levels (Oxenham & Dau, 2004; Summers, Boer, & Nutall, 2003; Summers & Leek, 1998), as well as in the presence of cochlear damage (Gifford et al., 2008; Recio & Rhode, 2000; Summers & Leek, 1998). These are conditions at which nonlinear active gain is reduced.

Due to its ability to measure an aspect of cochlear nonlinearity, Schroeder-phase masking has been used in many experiments and a number of clinical applications. For example, it has been shown to be more sensitive than conventional pure-tone audiometry in evaluating changes in nonlinear cochlear function resulting from damage during cochlear implant surgery (Gifford et al., 2008).

Previous psychophysical experiments involving Schroeder-phase masking have commonly used a three alternative forced choice (3AFC) method (Gifford et al., 2008; Kohrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2004; Summers & Leek, 1998). In that method, the participant chooses which of three randomly-ordered sets of Schroeder-phase masker stimuli also contains a probe tone. Most implementations of the procedure use a 3-down 1-up stepping rule to track a 79.4% correct level (Levitt, 1970). One disadvantage of the 3AFC method is its long testing duration. One participant was reported to require nearly 2 hours to be completely familiar with the task (Gifford et al., 2008; Oxenham & Dau, 2004). Our own data using this technique indicated that an average of 45 minutes was needed for a participant to complete the masking procedure for one frequency and intensity combination. The long testing duration may become intolerable to patients (especially the elderly) and make it unsuitable for clinical practice.

To address this issue, we developed a faster method of recording Schroeder-phase masking functions which may facilitate further studies using this technique. Our version of the Schroeder-phase masking procedure was inspired by the use of the Békésy tracking technique to record psychophysical tuning curves (PTCs) (Zwicker, 1974; Sęk et al., 2005). To record PTCs using this technique, the frequency of the masker is constantly swept and a repetitive probe of fixed frequency is presented at the same time. While the probe level is held constant, the masker level is increased and decreased according to the participant response – the masker level gradually increases when the participant holds down a button indicating that they can still hear the probe, and gradually decreases when they release it. To adapt this technique for Schroeder-phase masking, one of the authors (G.O'B.) developed software that slowly swept the scalar factor of a Schroeder-phase harmonic complex from -1.1 through 0 to +1.1, while a pulsatile probe tone was simultaneously presented and adjusted in level according to the participants' response (see Section 2.1.2.2 for more details). Using this new fast sweep method, the same masking function as that derived from the 3AFC method was able to be recorded in a much shorter time.

This paper presents preliminary findings on the reliability of the fast method in normal hearing subjects, namely the agreement of the new method as compared to the conventional method, and the repeatability of the new method (Chinn, 1990). Because various testing times can be predetermined using the Békésy tracking procedure by setting the sweep rate, we also address the question of how much faster the test can be without compromising its reliability. Three main aims of this paper are: i) to describe the development of the fast method of recording Schroeder-phase masking functions; ii) to study the reliability of the fast method by measuring both agreement and disagreement levels between the conventional and fast methods and the fast method's test-retest reliability; and iii) to establish the shortest testing time for the fast method without compromising its reliability.

2. Material and methods

2.1 The development of the fast method

The essential parameters for developing the fast method of recording Schroeder-phase masking functions will be described here. The important guidelines used for the conventional method (3AFC) will also be described here to allow comparison of the two methods. The parameters used for generating stimuli in both the conventional and fast methods were generally the same except for a few details which will be highlighted later. The main difference between the two methods is the Békésy threshold tracking procedure which greatly contributes to the shorter testing time of the fast method relative to the conventional method.

2.1.1 Stimuli

A pure tone was used as the probe, and the masker was a Schroeder-phase harmonic complex constructed from Equations 1 and 2. For the 3AFC technique, thresholds were obtained via adaptive procedures at nine discrete scalar factors (C = -1.00, -0.75, -0.50, -0.25, 0.00, 0.25, 0.50, 0.75 and 1.00). For the Békésy technique, the masker was swept from C = -1.1 to C = +1.1 in around 500 steps of 480 ms duration. Masker bandwidth ranged between 0.4 - 1.6 times the centre frequency (f_c) of the probe. This masker bandwidth was chosen on the basis that masker components ranging from $0.4 f_c - 1.6\ f_c$ were proven to effectively contribute to the masked threshold in Schroeder-phase masking (Oxenham & Dau, 2001a). The masker duration was set to be 480 ms with a 15 ms rise-fall time. The overall level of the masker was constant throughout the testing. The fundamental frequency (f_0) of the masker was varied according to the chosen number of masker components (n) based on this equation:

$$f_0 = (1.6 f_c - 0.4 f_c)/(n-1)$$

(Equation 3)

The probe was an intermittent pure tone (240 ms duration, 15 ms rise-fall time) presented simultaneously with the masker.

2.1.2 *Threshold tracking method*

Two threshold tracking methods are described below; the 3AFC technique (the conventional method) and the Békésy tracking technique (the fast method).

2.1.2.1 Three alternative forced choice method (3AFC)

The masker level was set at a constant level across the entire test. The probe level began at 10 dB higher than the masker level, and was altered automatically by custom-written software according to the participants' responses. As described in Figure 2 A, three masker stimuli of 480 ms duration (15 ms rise-fall time) were presented with a gap of 400 ms between each. One of the stimuli contained a 240 ms probe tone (15 ms rise-fall time) centrally embedded within the masker. Simultaneous with the auditory presentation, three buttons on the screen were highlighted sequentially, and the participant was asked to select the one that corresponded to the masker/probe combination. There was pause of 750 ms between the participants' selection and the next presentation. The probe level was decreased by 5 dB with every correct response until the participant responded incorrectly, at which point the probe level was increased by 10 dB until the participant started to respond correctly again. A change from responding correctly to incorrectly (or vice versa) was called a reversal. After the second reversal, the probe level was decreased by 1 dB for every correct response and increased by 2 dB for every incorrect response. The 1:2 ratio of the size of the downward and upward steps (S_{down} and S_{up} respectively) in both these cases was chosen according to the weighted up/down staircase (WUDR) adaptive algorithm of Kaernbach (1991), which defines the equilibrium conditions for convergence on a point (p) on the psychometric function as follows:

$$S_{down}\, p = S_{up}(1 - p)$$

(Equation 4)

As the chance level in a 3AFC test is 33.3%, a ratio of S_{down} to S_{up} of 1:2 was used to converge on a p of 66.7%, which represents the midpoint of the psychometric function. A total of 8 reversals were needed to complete a run. The first two reversals were discarded, and the response level of the remaining 6 reversal points was then averaged and taken as the threshold of the run. An average of at least two runs was needed to obtain the threshold at each of the scalar factors. If the difference in calculated threshold between two runs was more than 3 dB, an additional run was required and the threshold was then taken from the average of the two runs with the closest thresholds. All of the above steps were repeated for all 9 scalar factors to yield a complete Schroeder-phase masking function for one frequency/intensity combination (see Figure 2 A - C for an overview of the threshold tracking procedure for 3AFC using a 75 dB A masker at 500 Hz). This process took about 45 minutes. Participants were given a short training session prior to the actual testing.

2.1.2.2 Békésy technique

For the Békésy tracking technique, the masker was presented at a constant level and the probe was presented simultaneously, with the level being varied according to the participant's response. Initially, the probe level was 10 dB higher than the masker level, and participants were asked to respond by pressing a button as long as they could hear the probe, and release it as soon as they could not hear the probe. Visual feedback regarding the participants' responses was given whenever they were holding down the response button. The probe level was constantly decreased at a rate of 2 dB/s when the participant responded and constantly increased at this same rate when the participant did not respond. The transition from 'response' to 'non-response' or vice versa was taken as a 'reversal'. To ensure that the masked threshold at either -1.0 or +1.0 was correct, the sweep range of the scalar factor actually began at either -1.1 or +1.1 and did not commence sweeping until 2 reversals had been obtained at this starting scalar factor. The sweep rate was determined by the total scalar factor range (e.g. C = -1.1 to C = +1.1 = 2.2) divided by the chosen time for one sweep (e.g. 4 minutes or 240 s). For this scalar factor range, a 300 s sweep gave a rate of 0.0073 C/s, 240 s gave 0.00917 C/s, 180 s gave 0.0122 C/s, and so on. The stimuli were synthesised in blocks of 480 ms, with the scalar factor changing slightly between each block. To minimise any audible artefacts at the junction between these blocks, the first 5 ms of each stimulus was cross-faded with the final 5 ms of the previous one, after shifting the waveform so it gave the minimum error between the two cross-faded portions. The probe tone was also shifted by this same amount so that it did not shift relative to the masker.

To account for a known 'hysteresis effect' in Békésy tracking where measured curves can be shifted by small delays in participant responses (Sęk, Moore, Kluk, Wicher, 2005; Sęk & Moore, 2011), two sweep directions were applied: forward (from C = -1.1 to C = +1.1) and backward (from C = +1.1 to C = -1.1), with the final curve being the average of the forward and backward sweeps.

Figure 2 E) shows an example of the threshold tracking for one sweep of the fast method for a 500 Hz masker at 75 dB A. The change of state between 'response' and 'no-response' causes the decrease and increase of probe level throughout the sweep, which creates the ‘zig zag’ traces shown in the figure. The threshold is taken as the midpoint of the Békésy extrema. For example, in Figure 2 E) a ‘response’ at 52.8 dB A at C = -0.44 and a ‘no-response’ at 48.9 dB A at C = -0.45 are averaged to a threshold of 50.9 dB A at C = -0.445. Thresholds were calculated in this way at many points between C = -1.1 and C = +1.1, but linear interpolation was also used to derive thresholds at nine specific scalar factors that have been used in many previous studies of Schroeder-phase masking investigating cochlear function (C = -1.0 to +1.0 in increments of 0.25). The final Schroeder-phase masking function was obtained by averaging the masked thresholds at these points for the forward and backward sweeps.

2.1.3 Calibration

The custom software implemented an inverse filter process to compensate for the frequency response of the external sound card and the Sennheiser HD 280 Pro supra-aural headphones used in this part of the study. The frequency responses of the sound card and headphone were measured using a Brüel & Kjær Type 4128 Head and Torso Simulator (HATS) connected to a Brüel & Kjær 7539 5/1-ch. Input/Output Controller Module. The inverse filter process enabled computational estimates of sound levels to be made for each combination of stimulus, sound card and headphone, and ensured that the output level of each presentation was kept constant throughout testing.

2.1.4 Output measures obtained from the Schroeder-phase masking test

Figure 3 A shows examples of Schroeder-phase masking functions obtained at 75 dB A masker level and 500 Hz probe centre frequency using the conventional and fast methods. Both display the concave-up appearance typically recorded in healthy cochleae. As discussed above, the phase effect was calculated as the difference between the maximum and minimum masked thresholds in the Schroeder-phase masking function. The magnitude of the phase effect is related to the nonlinearity of cochlear function, with larger phase effects being observed in normal and healthy cochleae and reduced phase effects being observed in people with sensorineural hearing loss (Gifford et al., 2008; Kohlrausch & Sander, 1995; Lentz & Leek, 2001; Oxenham & Dau, 2004; Summers, 2000).

Another important measure is the location of the minima of the Schroeder-phase masking functions (for example, at C = +0.25 in Figures 3 C and 3 F). Previous studies have used the locations of the minima in the Schroeder-phase masking function to derive an estimation of the phase curvature of the auditory filter (Oxenham & Dau, 2001b; Oxenham & Ewert, 2005), as it is equal but opposite in phase to the Schroeder masker phase curvature producing the minimum threshold (Oxenham & Dau, 2001b)

2.2 Methods for experiment 1: Agreement of the fast method and the conventional method

One of the components in establishing reliability is the measure of agreement. This experiment was conducted to measure the agreement between the conventional and fast methods. The aim was to observe whether the fast method produced a result that agreed with the conventional 3AFC method.

2.2.1 Participants

38 normal hearing participants (aged 17-48 years old) participated in this study. All participants had normal audiometric thresholds of ≤ 20 dB HL at all octave frequencies, and had normal middle ear function as indicated by type A tympanometry. This study was conducted in sound treated booths at the University of Canterbury, New Zealand, and at the International Islamic University, Malaysia, and ethical approval was received from both institutions. All participants received a small honorarium for their participation.

2.2.2 Methods

Informed consent was obtained from participants prior to the testing. Preliminary hearing tests (otoscopy, tympanometry and pure-tone audiometry) were conducted prior to the Schroeder-phase masking tests to ensure that all participants had normal middle ear function and normal hearing thresholds between 250 Hz and 8 kHz. Each participant underwent Schroeder-phase masking tasks using both conventional (3AFC) and fast (Békésy) methods within one session. Practice was given prior to the actual testing for both conventional and fast methods. Participants were given the freedom to pause the test whenever they felt tired and to resume after taking a break. The order of the tests was alternated between participants.

All stimuli were presented monaurally through Sennheiser 280 Pro headphones via an external sound card. The masker consisted of 25 frequency components with a fundamental frequency of 25 Hz centred around 500 Hz (a range of 200 Hz – 800 Hz), as shown in Figure 1. The overall level of the masker was fixed at 75 dB A. The total duration of the masker was 480 ms (15 ms rise-fall time). The probe was a 240 ms 500 Hz tone (15 ms rise-fall time) used once per presentation in the conventional 3AFC method and repetitively in the fast method, as shown in Figures 2 A and 2 D.

The threshold tracking procedures described in Sections 2.1.2.1 and 2.1.2.2 were applied for the conventional 3AFC method and the fast method respectively. The time taken to complete either the forward or backward sweep was set at 5 minutes (300 sec) which gave a total testing time of 10 minutes for the fast method.

2.3 Methods for experiment 2: Test-retest reliability

25 participants from experiment 1 were asked to repeat the fast method to examine the consistency of the results between two trials within the same session. All the parameters for the fast method were kept the same as in experiment 1.

2.4 Method for *experiment 3: Determining the shortest testing time for fast method*

As the 10 minute testing time used in experiments 1 and 2 (5 minutes per sweep direction) was chosen somewhat arbitrarily, we decided to test if it was possible to shorten the testing time even more without compromising the reliability of the fast method. Six normal hearing participants from experiment 1 were asked to participate in this experiment. The total testing time (the sum of the forward + backward sweep durations) was set at 6 minutes, 8 minutes and 10 minutes. Other than the testing time (and therefore scalar factor sweep rate), all parameters for the fast method were kept the same as in experiment 1.

3. Results

3.1 Results of experiment 1

Figure 3 A shows the mean Schroeder-phase masking function obtained from both conventional and fast methods. Note that at each scalar factor the masked thresholds for both methods lie close to each other. The difference in masked thresholds between the conventional and the fast methods obtained for the same participant was calculated, and the mean difference across all participants at each scalar factor was plotted as in Figure 3 B. The mean difference was approximately zero dB across all scalar factors, and ranged from -2.2 dB (at C = -0.5) to 1.18 dB (at C = -1).

The difference in the magnitude of the phase effect between the conventional and fast methods was also calculated as a measure to assess the agreement between them. The intraclass correlation coefficient (ICC; Bartko, 1966) was used to analyze the agreement between the two methods (with a value of 1 showing perfect agreement and 0 showing no agreement), and was calculated using a two-way mixed ANOVA model in SPSS (version 20). The participant was treated as a random effect and the method was treated as the fixed effect (for a detailed explanation, see McGraw & Wong, 1996). Analysis gave an ICC value of 0.548 with a 95% confidence interval of 0.191 to 0.758. According to the criteria of Fleiss (1981), this ICC value indicated “fair-to-good” reliability.

The Information Based Measure of Disagreement (IBMD) was also calculated to measure the disagreement level between phase effects measured using the two methods, where 0 shows no disagreement and 1 shows total disagreement (Costa-Santos, Antunes, Souto & Bernardes, 2010). The IBMD value was calculated using an online calculator (Costa-Santos, 2012), and the result of 0.16 (with a 95% confidence interval of 0.133 to 0.182) showed low disagreement between the 2 methods. These two measures are discussed further later.

3.2 Results of experiment 2

The Schroeder-phase masking functions were recorded twice for the same participants within the same session using the fast method. The mean masked threshold (± one standard deviation) for each trial was plotted across all scalar factors, as shown in Figure 4 A. The difference in masked thresholds between the first and second trial at different scalar factors was calculated and the mean difference is plotted in Figure 4 B. Note that the mean difference between the masked threshold between first and second trial lay at approximately 0 dB across all scalar factors, indicating high repeatability. In measuring the test retest reliability, the consistency of the phase effect measured from 2 trials of the fast method of Schroeder-phase masking was analyzed using a one-way random ANOVA in ICC analysis (see McGraw & Wong, 1996). Analysis gave the ICC value of 0.773 with 95% confidence interval between 0.554 and 0.893.

3.3 Results of experiment 3

The results for different fast method testing times were compared to those for the conventional 3AFC method. Results for the conventional method for these 6 participants were taken from experiment 1 and served as the reference for the comparison. The masked threshold at each scalar factor was taken from the average of 6 participants.

Figure 5 A shows the plot of the Schroeder-phase masking functions for different fast method testing times (6, 8 and 10 minutes) and for the 3AFC method (45 minutes on average). The level of the masked threshold at the minima of the Schroeder-phase masking functions decreased with increasing testing time. This resulted in a larger phase effect being observed with longer testing times, as plotted in Figure 5 B; with the largest phase effect being observed for the conventional method followed by the fast method at 10 minutes, 8 minutes and 6 minutes. The comparisons of phase effect between different testing times for the fast method and the conventional method were conducted using linear mixed effect models (testing time as the fixed effect and the participant as the random effect) in SPSS (version 20). A significant difference in the phase effect was observed between the conventional method and the 6-minute fast method ($p = 0.043$) - refer to Figure 5 B. However, comparison of the conventional and fast methods at 8 minutes and 10 minutes testing times gave no significant difference ($p = 0.120$ and 0.243 respectively), suggesting that the testing time of the fast method can be shortened to 8 minutes before it starts to give results that are different from the conventional method.

4 Discussion:

While previous studies have used a variety of different test parameters that made direct comparisons with our results difficult, we were able to compare our results with those of Gifford et al. (2008) due to the similar parameters used in our studies. Using a similar fundamental frequency, masker bandwidth and masker intensity to those used in our study, Gifford et al. (2008) measured a phase effect of around ≈20 dB in normal hearing participants, consistent with our findings. Our findings also showed that the average minimum of the masking function occurred at $C = 0.25$ (Figure 3 A, consistent with that in Gifford, et al. (2008). A minimum that lies at positive ‘C’ was consistent with the notion that the location of minima reflects the phase curvature of the auditory filter in an opposite manner, indicating that it has negative phase curvature (Kohlrausch & Sander, 1995; Oxenham & Dau, 2001b).

In establishing the reliability of a newly developed method, there are no fixed guidelines as to which statistical analysis should be used to measure agreement between these two methods.

Several types of analysis have been proposed previously, each with its own limitations (Luiz & Szklo, 2005). The use of Intra Class Correlation (ICC) analysis (Bartko, 1966) in measuring the reliability of medical instruments was the most commonly reported in previous studies, followed by the mean comparison method and the Bland-Altman method (Zaki, Bulgiba, Nordin & Ismail, 2013). Rankin & Stokes (1998) proposed the use of ICC together with Bland and Altman's (1986) method for reliability assessment. However, the Bland-Altman method requires knowledge of decision rules or clinically accepted values of how far apart the two measurements can be without causing difficulties. This key requirement has been viewed as a weakness of the Bland-Altman method, since determining the decision rules is not a straightforward process if the conventional technique is not being set as the 'gold standard' (Alanen, 2010). This holds true at least for this study, as the lack of knowledge on determining the clinically accepted decision rules makes the Bland-Altman analysis inappropriate for our data: To the best of our knowledge, there is no data showing how much the phase effect or masked threshold can differ between two measurements before it starts to give a clinically significant effect, and so for our agreement analysis, we did not intend to compare our results with any clinically significant value.

Acknowledging there are specific advantages and disadvantages of every analysis method, Luiz & Szklo (2005) proposed the use of more than one analysis to measure the agreement between two methods. As per the recommendations of Luiz & Szklo (2005), ICC analysis (from the two-way model) (Bartko, 1966), Information Based Measure of Disagreement (Costa-Santos, Antunes, Souto, & Bernardes, 2010), and mean comparison (mixed linear model analysis) were used to measure the agreement and disagreement levels between the fast and conventional methods respectively in our study.

ICC is a measure of the proportion of a variance that is attributable to the measure of interest, which gives information about the measurement error of the measurement of interest, and has been widely used as a reliability index (Shrout & Fleiss, 1979). The relatively low ICC of agreement value obtained in this study (0.548, indicating "fair-to-good" reliability) might be due to the homogeneity of the population (i.e. the participants all had normal hearing). One important limitation of ICC is that it is strongly influenced by the variance of the traits in the population in which it assessed (Müller & Büttner, 1994). A homogenous population such as ours with low variance may result in a low ICC.

The Information Based Measure of Disagreement (IBMD) was also calculated (Costa-Santos et al., 2010). The IBMD is a metric system that was calculated based on the amount of information obtained from the differences between 2 variables (Costa-Santos et al., 2010). The IBMD gives a normalized value ranging from 0 to 1, with '0' showing no information in the differences between 2 variables. The IBMD value increases when the difference between 2 variables increases, tending to 1. The IBMD value of 0.16 shows a low level of disagreement between phase effects measured using conventional and fast methods. Furthermore, mean comparison analysis using the mixed linear model in Experiment 3 showed no significant difference in phase effect between the conventional and fast methods (for 8 and 10 minutes testing time). Indeed, the absolute mean differences in masked thresholds between the 2 methods lies at approximately 0 dB (ranging from -2.2 dB at C = -0.5 to 1.18 dB at C = -1). The acceptable level of 'ICC of agreement' value, low IBMD value, and no significant difference in phase effect found between the conventional and the fast method, supports the agreement of the fast method with the conventional method.

Internal consistency, also known as test-retest reliability or repeatability is another important aspect to focus on in establishing the reliability of a new method. Repeatability can be analyzed using the one-way ANOVA model of ICC. A different ranking system was proposed by Nunnally, Bernstein, & Berge (1967) for repeatability analysis, with ICC of <0.70 being unacceptable, 0.70 – 0.8 considered fair and acceptable, and >0.80 considered excellent. Our ICC value for the repeatability analysis was 0.773 and fell within the acceptable limits. The reliability of the fast method is also supported by the mean difference in masked thresholds between the first and second trial laying at apporoximately 0 dB, as shown in Figure 4 B.

The main difference between the conventional and the fast methods of Schroeder-phase masking lies in the threshold tracking procedures and the psychophysical method of threshold seeking. Apart from this, all other technical parameters for testing were kept the same (i.e. the probe duration and level, the masker level, the sound card, calibration method, transducer, testing environment etc.). With these technical parameters being kept constant for both the conventional 3AFC and fast method, any difference in terms of perception should be reflected primarily in the participant reliability, which can be affected by factors such as their individual motivation and decision levels. Several studies on the reliability of hearing threshold determination showed that there were no differences in intra-subject reliability across different tested frequencies (Henry, Flick, Gilbert, Ellingson, & Fausti, 2001; Mahomed, Eikelboom & Soer, 2013; Sinks & Goebel, 1994). If there were any difference in the participant's perception as a result of applying two different psychophysical methods, we would expect the difference (measured as threshold) to have manifested at one frequency as much as at any other. Based on this reasoning, testing at one frequency (e.g. 500 Hz, as we have here) should provide a reasonable reliability measure for the fast method.

In conclusion, this paper presented preliminary findings on several aspects of the reliability of a fast method of Schroeder-phase masking. Having demonstrated two important properties (agreement with the conventional method and repeatability), the fast method of Schroeder-phase masking has been shown to be a reliable technique. The main strength of this method is the 80% reduction in testing time it gives relative to the conventional method (8 minutes compared to 45 minutes), which may considerably facilitate future research using Schroeder-phase masking. Further studies on hearing-impaired participants are needed to establish the validity of the fast method, and these are currently in progress.

Acknowledgements:

The authors would like to thank Prof. Don Sinex for helpful discussions. Thanks to the Chair of Malay Studies grant (Victoria University of Wellington), GN ReSound NZ Ltd. (the UC ReSound ReSearch Fund), IIUM research funding, and the Ministry of Education Malaysia for funding data collection. Thanks also to the Inner Ear Biology committee and the Ministry of Education for funding the travel to the IEB 2014 conference in Kyoto, Japan, where part of this study was presented.

References:

Alanen, E. (2010). Everything all right in method comparison studies? *Statistical Methods in Medical Research, 21*(4), 297-309.

Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. Psychological Reports, 19, 3-11.

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The lancet, 327(8476), 307-310.Carlyon, R. P., & Datta, A. J. (1997). Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system. *Acoustical Society of America, 101*(6), 3636-3647.

Chinn, S. (1990). The assessment of methods of measurement. *Statistics in Medicine, 9*, 351-362.

Costa-Santos, C., Antunes, L., Souto, A., & Bernardes, J. (2010). Assessment of disagreement: a new information-based approach. *Annals of epidemiology, 20*(7), 555-561.

Costa Santos, C. (2012). *Information Based Measure of Disagreement On-line Calculator [Computer software].* Retrieved from http://disagreement.med.up.pt

Fleiss J.L. (1981). Statistical methods for rates and proportions. New York: Wiley

Gifford, R. H., Dorman, M. F., Spahr, A. J., Bacon, S. P., Skarzynski, H., & Lorens, A. (2008). Hearing preservation surgery: Psychophysical estimates of cochlear damage in recipients of a short electrode array. *The Journal of the Acoustical Society of America, 124*(4), 2164. doi: 10.1121/1.2967842

Henry, J. A., Flick, C. L., Gilbert, A., Ellingson, R. M., & Fausti, S. A. (2001). Reliability of hearing thresholds: computer-automated testing with ER-4B Canal PhoneTM earphones. Journal of rehabilitation research and development, 38(5), 567-581.

Kaernbach, C. (1991). Simple adaptive testing with a weighted up-down method. *Perception & Psychophysics, 49,* 227-229.

Kohlrausch, A., & Sander, A. (1995). Phase effects in masking related to dispersion in the inner ear.II. Masking period pattern of short targets. *Journal of Acoustical Society of America, 97*(3), 1817-1829.

Lentz, J. J., & Leek, M. R. (2001). Psychophysical Estimates of Cochlear Phase Response: Masking by Harmonic Complexes. *JARO - Journal of the Association for Research in Otolaryngology, 2*(4), 408-422. doi: 10.1007/s101620010045

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*, 467-477.

Luiz, R. R., & Szklo, M. (2005). More than one statistical strategy to assess agreement of quantitative measurements may usefully be reported. *Journal of clinical epidemiology*, *58*(3), 215-216.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.

Mahomed, F., Eikelboom, R. H., & Soer, M. (2013). Validity of automated threshold audiometry: a systematic review and meta-analysis. Ear and hearing,34 (6), 745-752.

Müller, R., & Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in medicine*, *13*(23-24), 2465-2476.

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. T. (1967). *Psychometric theory* (Vol. 226). New York: McGraw-Hill.

Oxenham, A. J., & Dau, T. (2001a). Reconciling frequency selectivity and phase effects in masking. *J Acoust Soc Am., 110*(3), 1525-1537.

Oxenham, A. J., & Dau, T. (2001b). Towards a measure of auditory-filter phase response. *J Acoust Soc Am., 110*, 3169-3178.

Oxenham, A. J., & Dau, T. (2004). Masker phase effects in normal-hearing and hearing-impaired listeners: Evidence for peripheral compression at low signal frequencies. *The Journal of the Acoustical Society of America, 116*(4), 2248. doi: 10.1121/1.1786852

Oxenham, A. J., & Ewert, S. D. (2005). Estimates of auditory filter phase response at and below charachteristic frequency J Acoust Soc Am., 117(4), 1713-1716.

Patterson, R. D., & Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representation of frequency resolution," in "Frequency selectivity in hearing", edited by B. C. J. Moore, Academic Press.

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clinical rehabilitation, 12, 187-199. Recio, A., & Rhode, W. S. (2000). Basilar membrane responses to broadband stimuli. *J. Acoust. Soc. Am. , 108*(5), 2281-2298.

Schroeder, M. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.). *Information Theory, IEEE Transactions on, 16*(1), 85-89.

Sęk, A., Alcántara, J., Moore, B. C., Kluk, K., & Wicher, A. (2005). Development of a fast method for determining psychophysical tuning curves. *International Journal of Audiology*, *44*(7), 408-420.

Sek, A., & Moore, B. C. (2011). Implementation of a fast method for measuring psychophysical tuning curves. *International Journal of Audiology*, *50*(4), 237-242.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420-428.

Sinks, B. C., & Goebel, J. A. (1994). Test-Retest Reliability of High-Frequency Thresholds at Bedside with Sensorineural Hearing-Impaired Listeners. JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY, 5, 399-399.

Summers, V. (2000). Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes. *Journal of Acoustical Society of America, 108*(5), 2307-2317.

Summers, V., Boer, E. D., & Nuttall, A. L. (2003). Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects. *Acoustical Society of America, 114*(1), 294-306.

Summers, V., & Leek, M. R. (1998). Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. *Hearing Research, 118*, 139-150.

Zaki, R., Bulgiba, A., Nordin, N., & Ismail, N. A. (2013). A Systematic Review of Statistical Methods Used to Test for Reliability of Medical Instruments Measuring Continuous Variables. *Iran J Basic Med Sci, Vol., No. Jun 16*(6), 803-807.

Zwicker, E. (1974). On a psychoacoustical equivalent of tuning curves. *Facts and models in hearing* (pp. 132-141): Springer Berlin Heidelberg.

Figure 1:

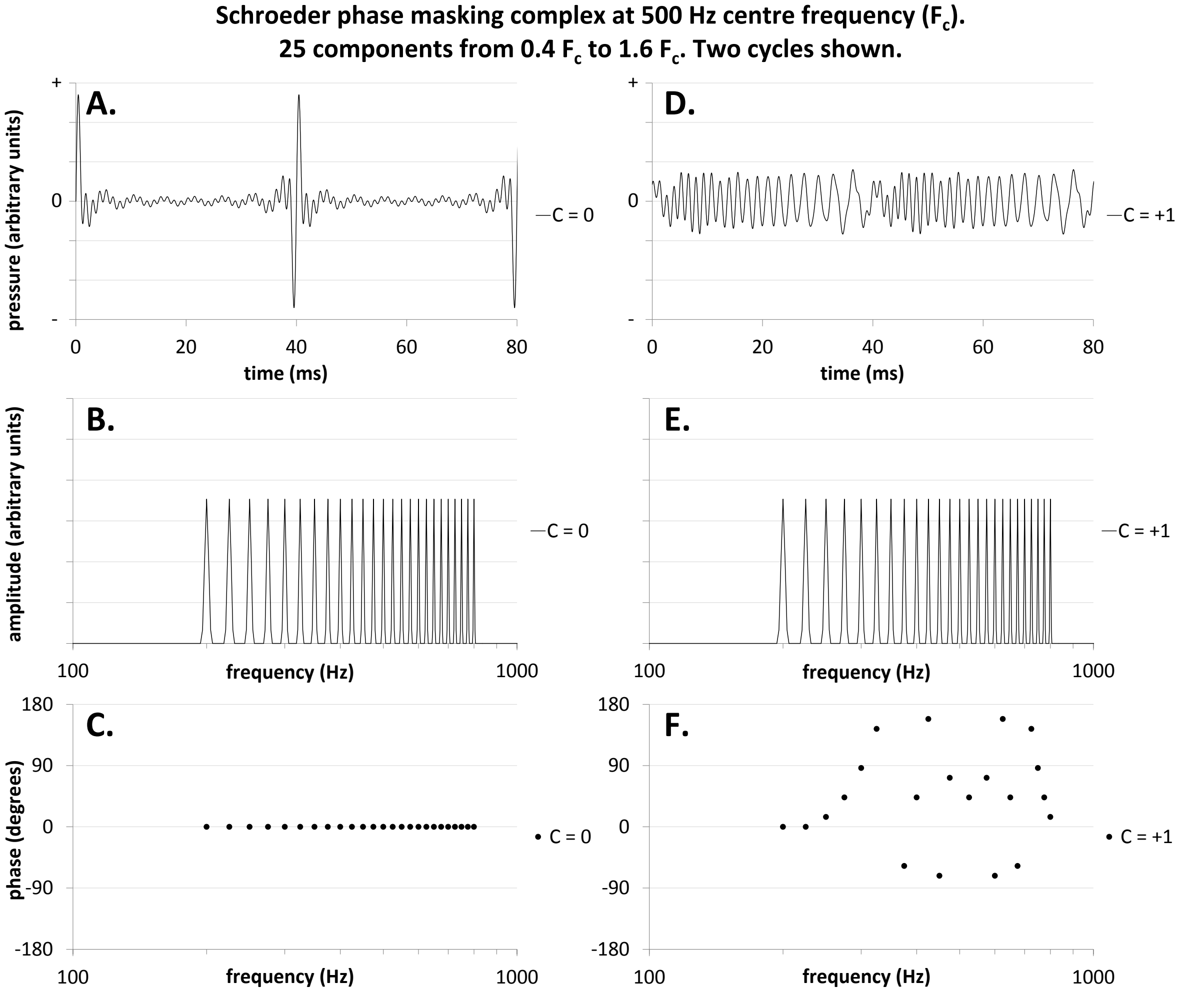

Figure 1 caption:

Illustration of the effect of Schroeder-phase settings on the crest factor of a harmonic complex consisting of 25 sinusoids ranging from 200 to 800 Hz at 25 Hz intervals. The left column (panels A, B, and C) shows two cycles of the time domain waveform, and the amplitude and phase spectra when the phases of all components are set to zero (Equation 2's scalar factor C = 0). The right column (panels D, E, and F) shows the masker when Schroeder-phases are used (scalar factor C = +1). Note the reduced crest factor in the time domain waveform (D), the unchanged amplitude spectrum (E), and the non-zero Schroeder-phase spectrum (F). When the scalar factor C = -1, the phase spectrum (F) is flipped vertically and the time domain waveform (D) is flipped horizontally.

Figure 2:

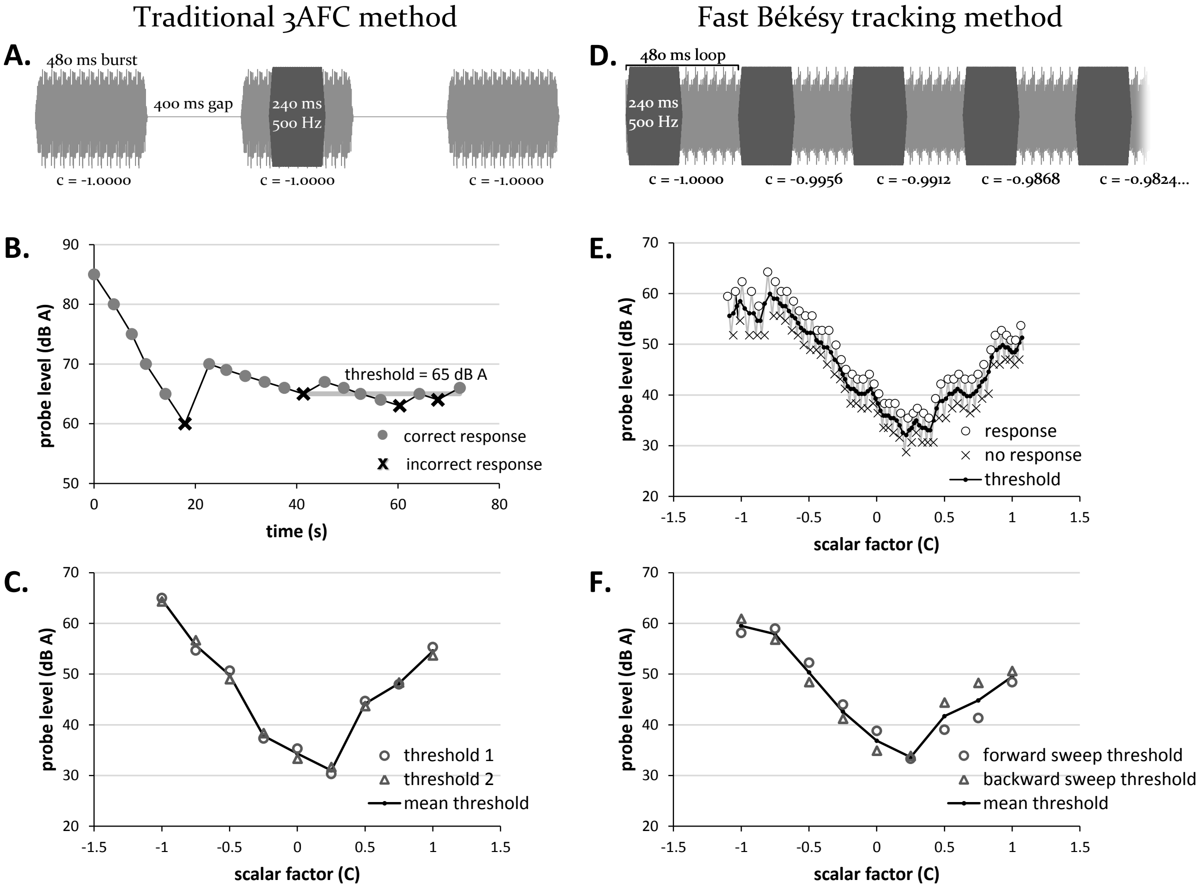

Figure 2 caption:

A comparison of the threshold tracking procedures used in the 3AFC and fast methods at 75 dB A masker level for 500 Hz. In the 3AFC procedure: a) the participant chooses which of the three masker intervals contains the probe tone; b) an adaptive algorithm alters the probe level, with around 25 presentations being required to obtain threshold for one run at each scalar factor; and c) the final threshold is the average of two runs. Total testing time = 45 minutes to record the curve. In the fast method: d) The participant holds a key while they can detect the probe in the presence of a slowly changing continuous masker; e) the probe level changes according to participant's response while the masker scalar factor is slowly swept from -1.1 to +1.1 over 4 or 5 minutes; and f) the final threshold is the average of the forward and backward sweeps. Total testing time = 8 minutes to record the curve.

Figure 3:

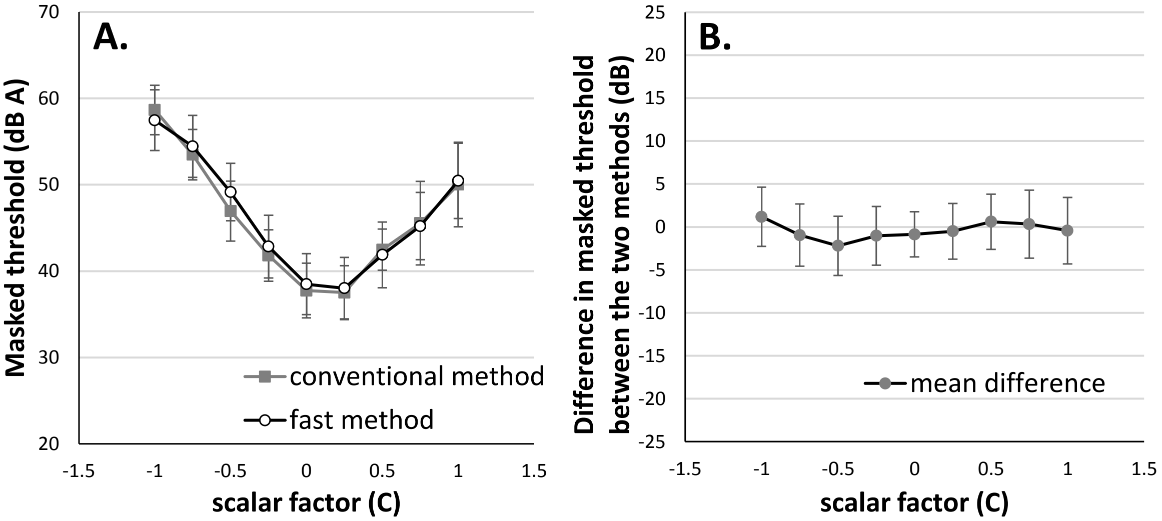

Figure 3 caption:

A) Average (n = 38) Schroeder-phase masking functions obtained using the conventional (square) and fast methods (circle). B) Difference in masked threshold between the conventional and fast methods at different scalar factors. Error bars indicate 1 standard deviation.

Figure 4:

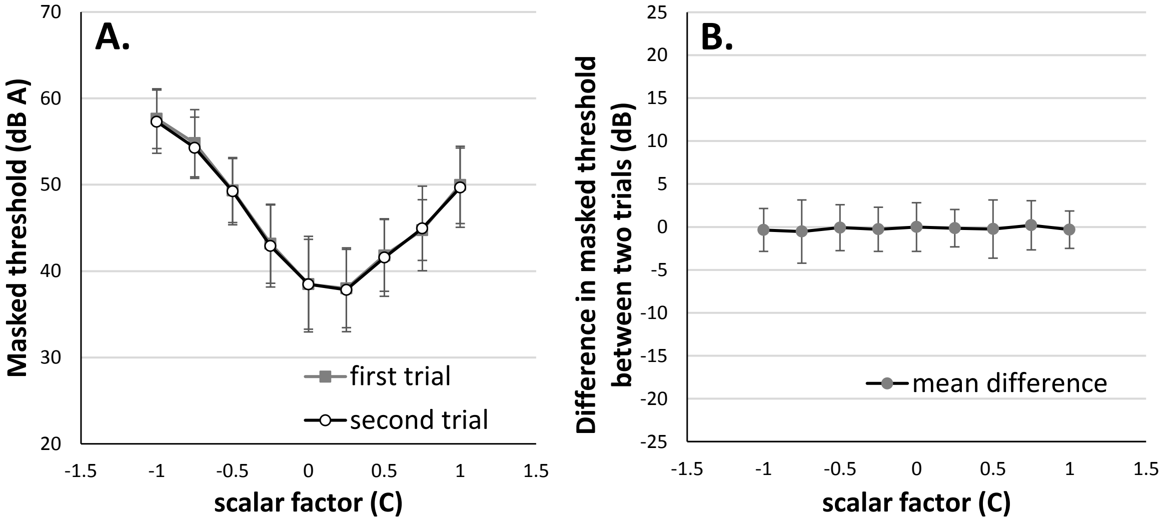

Figure 4 caption:

Average (n = 25) Schroeder-phase masking functions for repeated trials within the same session. B) Mean difference in masked threshold between the first and the second trial at different scalar factors. Error bars indicate 1 standard deviation.

Figure 5:

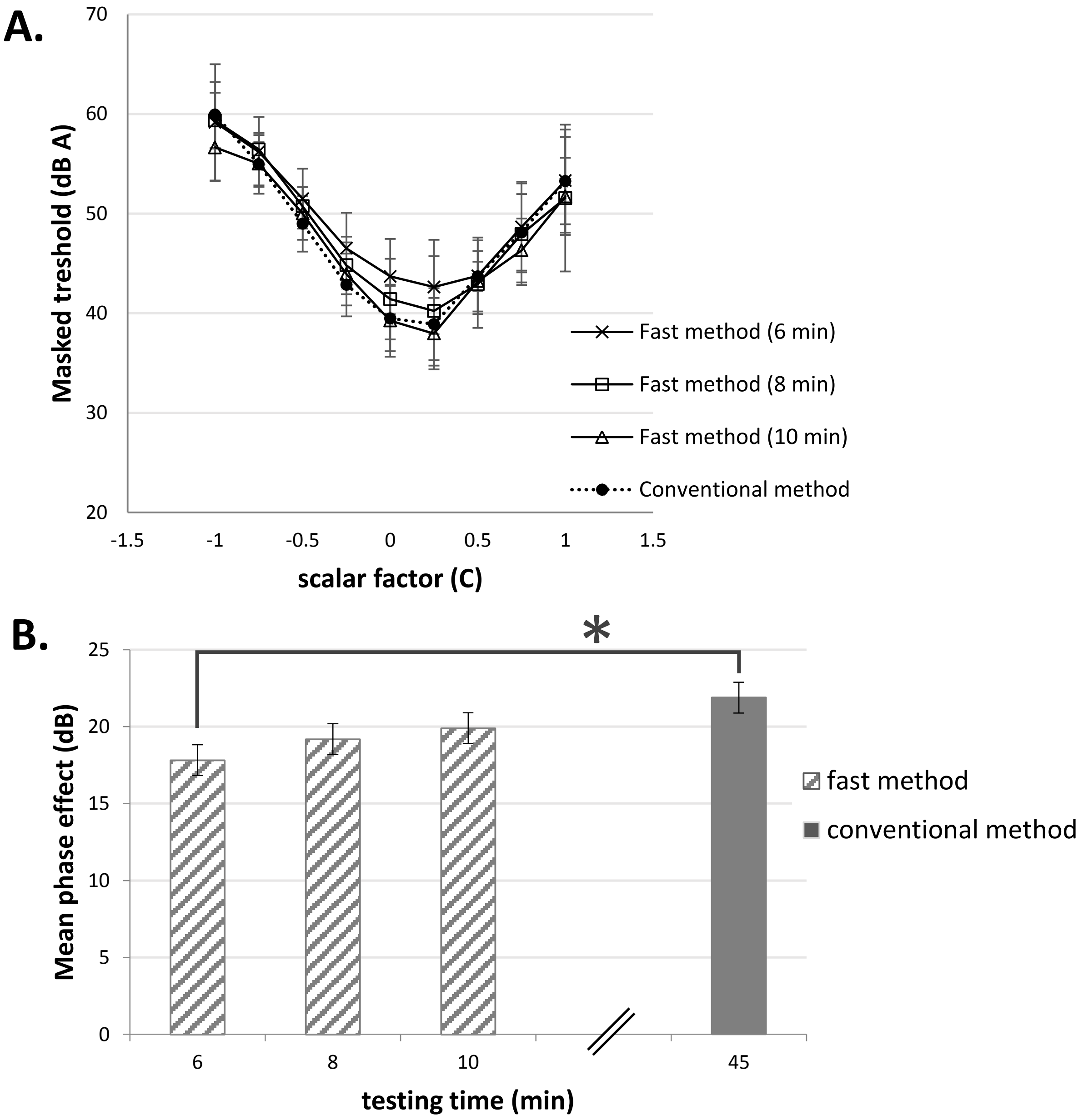

Figure 5 caption:

Average (n = 6) Schroeder-phase masking functions (A) and phase effect magnitudes (B) obtained with different fast method testing times (6, 8 and 10 minutes) and the conventional 3AFC method. Error bars show ± 1 standard deviation. The asterisk indicates a statistically significant difference (p = 0.043) between the 6 minute fast method and the conventional 45 minute 3AFC method.