Can place-specific cochlear dispersion be represented by auditory steady-state responses?

Paredes Gallardo, Andreu; Epp, Bastian; Dau, Torsten

Published in:
Hearing Research

Link to article, DOI:
10.1016/j.heares.2016.02.014

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Can place-specific cochlear dispersion be represented by auditory steady-state responses?

Andreu Paredes Gallardo, Bastian Epp, Torsten Dau*

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads Building 352, 2800 Kongens Lyngby, Denmark

1. Introduction

The frequency-place transformation that characterizes the mammalian auditory system is established in the inner ear, the cochlea, and is mainly a consequence of the mechanical properties of the basilar membrane (BM) as part of the cochlea, with an impedance that varies as a function of position (place) along the BM. The corresponding frequency analysis in the cochlea, associated with this frequency-place transformation, has traditionally been described effectively as a bank of overlapping band-pass filters (e.g., Fletcher, 1940). The properties of these filters in humans have been estimated mainly using psychoacoustic paradigms. Based on data from behavioral masking experiments, various studies have estimated the amplitude characteristics of the band-pass filters, commonly referred to as “auditory filters” (e.g., Fletcher, 1940; Zwicker et al., 1957; Patterson, 1976; Glasberg and Moore, 1990; Rosen and Baker, 1994). More recently, there has been increasing focus on describing the auditory filters’ phase response (e.g., Lentz and Leek, 2001; Oxenham and Dau, 2001; Shen and Lentz, 2009; Wojtczak and Oxenham, 2009). Such estimates of the auditory filter phase response have typically been obtained from masking experiments using tones presented in a particular type of harmonic tone complex maskers, known as Schroeder tone complexes.

Schroeder tone complexes consist of equal-amplitude tones with a starting phase $\theta_0$ which follows an algorithm developed by Schroeder (1970). The waveform of such tone complexes resembles a periodic series of linear downwards $(\text{Sch}+)\text{ or upwards (Sch}-)$ frequency sweeps, depending on the specific phase configuration. $\text{Sch}+$ and $\text{Sch}-$ stimuli thus have identical power spectra, but differ in their phase spectra. Masked thresholds of tonal signals that are spectroly centered in the Schroeder tone complex have been found to be $5-25$ dB lower in the $\text{Sch}+$ masker condition than in the $\text{Sch}$-masker condition (e.g., Smith et al., 1986;
Kohrausch and Sander, 1995), demonstrating that concepts based
on the power of the stimuli, such as the power spectrum model of
masking (e.g., Fletcher, 1940; Moore, 2003), fail to predict the
perceptual results in these conditions. Instead, the threshold dif-
fences obtained with the Sch-vs. Sch+ maskers have been
explored by analyzing the interaction between the stimulus phase
spectrum and the phase transfer function of the auditory filter.
If the phase curvature of an auditory filter, i.e. the second derivative
of the corresponding phase transfer function, would be of the
same size, but opposite in sign, as the phase curvature of the Sch+
masker, this should result in a highly modulated internal repre-
sentation. In contrast, conceptually, the Sch-masker should lead to
a flat, or less peaky, internal representation after the trans-
formation through the same auditory filter. Thus, the peaky in-
ternal representation obtained in response to the Sch− stimulus
should be less effective as a masker than the Sch-stimulus, since
the signal can be more easily detected in the dips of the peaky
representation (e.g., Kohrausch and Sander, 1995).

Following the hypothesis that minimum masking is achieved
when the phase curvature of the stimulus is equal in magnitude
but opposite in sign to that of the auditory-filter centered at the
signal frequency, Oxenham and Dau (2001b) behaviorally esti-
mated and Dau (2001b) the auditory filter phase curvature for a broad range of
center frequencies (fc). However, it remained unclear to what
extent effects of cochlear compression, level-dependent auditory
filter bandwidth as well as properties of temporal envelope pro-
cessing contribute to the observed masked patterns both in
normal-hearing and hearing-impaired listeners (Oxenham and
Dau, 2001a). Furthermore, such studies have assumed a roughly
constant auditory-filter phase curvature within the auditory-filter’s passband (e.g., Oxenham and Dau, 2001b; Kohrausch and
Sander, 1995; Lentz and Leek, 2001), which may only be a valid
approximation in the frequency region around the center fre-
cency but not throughout the entire passband of the filter (e.g.,
Shen and Lentz, 2009). Finally, in the psychophysical studies
investigating auditory filter phase response (e.g., Lentz and Leek,
2001; Oxenham and Dau, 2001b; Shen and Lentz, 2009;
Wojtczak and Oxenham, 2009) only the output from the audi-
tory filter providing the highest signal-to-noise ratio (SNR) has
typically been considered in the detection process (Moore, 2003).
Thus, the correspondence of such behavioral estimates with the
properties of the cochlear filters may be only qualitative and
direct.

Physiological responses would provide more detailed and local
information about cochlear filter phase response. However, these
are invasive and can only be obtained in animal studies. Indeed,
physiological measures of BM responses to Schroeder tone com-
plexes have been performed in chinchillas (Reco and Rhode,
2000) and guinea pigs (Summers et al., 2003). Both studies
concluded that BM responses to Sch− stimuli are “peakier” than
the responses to Sch+ stimuli, qualitatively consistent with the
results from the psychophysical studies in humans described
above. However, the connection between such physiological re-
cordings in animals and the human cochlear phase response re-
mains indirect.

In the present study, human auditory evoked potentials (AEP)
were considered to investigate effects of local cochlear dispersion.
In contrast to earlier studies that used broadband stimuli designed
to compensate for “spatial dispersion” (i.e., travel-time differences
across frequency) and, as a consequence, to create maximal neural
synchronization across the cochlear partition (e.g., Dau et al., 2000;
Junius and Dau, 2005; Elberling et al., 2007), the approach taken in
the present study was complementary in the sense that bandlimited
stimuli that were designed to compensate for “temporal disper-
sion” of the response at a restricted segment on the cochlear
partition were used. It has been demonstrated that AEPs, such as
auditory steady state responses (ASSR) obtained with narrowband
stimuli (like amplitude modulated tones) can provide reliable and
frequency-specific information, e.g. in connection to objective
auditory threshold estimations (e.g. Picton, 2011). Here it was
tested if properties of local cochlear dispersion can be represented
by ASSRs obtained with an “optimized” (bandlimited) stimulation.
Two stimulation paradigms were chosen: One of them considered
Schroeder tone complexes as in the studies described above; the
other one used narrowband stimuli, built from estimates of
cochlear filter impulse responses (IR) as proposed by Irino and
Patterson (2001). The IRs models have been argued to account for
psychophysical masking data in humans (Rosen and Baker, 1994) as
well as for auditory-nerve responses obtained from physiological
measurements in cats (Carney et al., 1999). Here, the use of the
temporally reversed IR as stimulus was hypothesized to theoreti-
cally produce a maximally peaky response at the output of the
corresponding filter.

2. Method

2.1. Stimuli

Signal generation and data post-processing routines were
implemented in Matlab, and the utility Playrec was used to manage
the communication between Matlab and the soundcard. The digi-
tal-to-analog (D/A) conversion of the acoustic stimulus was per-
formed at a sampling frequency, fs, of 48 kHz with a 24-bit
resolution.

2.1.1. Schroeder tone complexes

Schroeder tone complexes were generated by adding harmonics
between 0.4 and 1.6 times the center frequency (fc), following the
study of Oxenham and Dau (Oxenham and Dau, 2001a). Sound
pressure levels (SPL) of 45 and 70 dB were considered. fc was set to
1 kHz and a fundamental frequency (f0) of 90 Hz was chosen since a
repetition rate of 90 Hz has been shown to produce strong ASSR
primarily generated in the auditory brainstem (e.g., Picton, 2011).
The initial phase of each harmonic was defined according to
equation (1.1), where N stands for the total number of harmonics
and C is a scalar, introduced by Lentz and Leek (Lentz and Leek,
2001) to allow sweep rate modifications.

\[
\theta_n = C \pi n \left(\frac{n - 1}{N}\right)
\]  

(1.1)

Three C values were tested for each level: C = +1 (corresponding
to the original Sch+ stimulus), C = 0 and C = −1 (corresponding
to the original Sch-stimulus). In the case of C = 0, all components
were added with zero phase, resulting in a series of frequency-limited
pulses. Schroeder tone complexes have a constant phase curva-
ture, being positive for C > 0, zero for C = 0 and negative for C < 0.
Thus, Sch− complexes represent periodic sequences of linear
downward frequency sweeps whereas Sch-tone complexes are
temporally reversed, i.e. upward frequency sweeps. The phase
curvature of a given Schroeder tone complex (Kohrausch and
Sander, 1995) amounts to:

\[
d\theta = 2\pi
\]

(1.2)

\[
d\theta
\]
2.1.2. Gammachirp impulse responses

Gammachirp impulse responses (gc-IR) were also used as stimuli, generated using the model of Irino and Patterson (Irino and Patterson, 2001) and including level-dependent cochlear compression. The gc-IR represent upward frequency sweeps. As in the case of the Schroeder tone complexes, the corresponding temporally reversed stimulus was also considered, i.e. the reversed impulse response, gc-IR\textsubscript{rev}, representing a downward frequency sweep. The IRs were obtained at \( f_c = 2\) kHz for 50 and 70 dB SPL and concatenated with a repetition rate of 80 Hz in order to obtain an integer number of samples per repetition.

2.2. ASSR recordings

2.2.1. Subjects

A total of 14 adult subjects participated in the experiments (9 males and 5 females), between 22 and 32 years of age (mean of 26 years). All subjects had normal hearing according to their pure-tone audiogram (\( \leq 20\) dB HL). The study was approved by the Danish Science-Ethics Committee (Den Nationale Videnskabsetiske Komité) under the reference H-3-2013-004.

2.2.2. Apparatus

ASSRs were recorded in a double-walled, sound-attenuating and electromagnetic shielded booth. The Biosemi ActiveTwo system, an electroencephalography (EEG) recording system with active electrodes, was used for the recording of the ASSRs. The signal was played to the subjects in epochs of 1 s through an earphone (ER-2 insert earphone from Etymotic Research Inc.).

The Biosemi ActiveTwo system was set at speed mode 6 with a decimation factor of 1/2, leading to a sampling frequency of 4096 Hz with 24-bit resolution. The hardware filter bandwidth follows a 5th order sinc response, with the half-power bandwidth, the quality (Q) factor, was set to be constant along the segments.

The electrodes were placed on the scalp by using an elastic cap with plastic electrode holders. The cap follows the standardized 10/20 electrode system with 64 electrode holders, but only signals from Cz (vertex electrode) and P10/P9 (behind the right/left ear), together with the CMS and DRL, were recorded. All ASSR results shown in the present study represent the Cz-P10 difference (right ear measurements) or the Cz-P9 difference (left ear measurement). The signals from the electrodes and the trigger were stored on a hard disk using the ActiView software.

2.2.3. Recording and analysis of ASSR

The recording sessions lasted about 2 h in the case of the Schroeder tone complexes and about 30 min in the case of the IR stimuli. The subjects were seated in a comfortable reclined chair and were encouraged to relax or sleep during the recordings.

The recorded EEG data were first filtered with a band-pass fifth order Butterworth filter (60–300 Hz) applied in both forward and backward directions. Epochs exceeding an amplitude threshold of \( \pm 80\) \( \mu \)V were rejected. The non-rejected epochs were concatenated forming sweeps of 16 epochs, resulting in a spectral resolution of 0.0625 Hz/bin. The resulting sweeps were averaged, ideally reducing the noise term by a factor of the square root of the number of averaged trials.

An F-test statistical method was applied to the filtered and averaged EEG spectrum in order to objectively assess ASSR significance. The power at the ASSR frequency was compared to the power of the EEG background noise, defined as the power average at \( \pm 3\) Hz (96 bins) from the ASSR frequency (F-ratio). At the ASSR bin, the power estimate represents a chi-squared variable with 2 degrees of freedom. In contrast, the power noise estimate is a chi-squared variable with 2–96 degrees of freedom (Dobie and Wilson, 1996). With the degrees of freedom of each variable and the F-ratio, the F cumulative distribution was calculated. The probability (p) of the ASSR power being at least as big as the current observation was calculated as a 1–F cumulative distribution function, assuming that both noise and ASSR component came from the same distribution. If the null hypothesis (p \( \leq 0.01 \)) was rejected, the ASSR component was considered statistically significant from the EEG background noise estimate (positive F-test).

A Kruskal-Wallis one-way analysis of variance was used to test whether the differences in ASSR magnitude for different conditions were statistically significant. A significance level of \( a = 0.01 \) was chosen. If p \( \leq 0.01 \), the null hypothesis was rejected in favor of the alternative hypothesis, implying that the results came from different distributions.

2.3. Simulation of frequency selectivity and cochlear dispersion

BM motion simulations were obtained using a one-dimensional, linear transmission line model to illustrate effects of cochlear dispersion. The linear and passive version of the model developed by Epp et al. (2010) was chosen, which discretizes the cochlea into 1000 equally spaced segments. The ratio between the resonance frequency of each segment and the half-power bandwidth, the quality (Q) factor, was set to be constant along the segments. Different Q-factors were used to simulate different stimulus levels to reflect level dependent BM tuning (Epp et al., 2010).

3. Results

3.1. ASSR recordings

Fig. 1 shows the ASSR magnitudes in response to the Schroeder tone complexes as a function of the C value. The results for the individual subjects are shown in the upper panels for the stimulation levels of 45 dB SPL (left) and 70 dB SPL (right). The gray scale of the symbols indicates the SNR of the response on a dB scale. The single open symbol indicates that the response in this condition was found to be statistically significant for condition C = 1.

Overall, the ASSR amplitude elicited with the Schroeder tone complexes was found to be significantly larger for C = 1 than for C = +1, both at 45 dB SPL (\( \chi^2(1) = 12.79, p < 0.001 \)) and at 70 dB SPL (\( \chi^2(1) = 14.29, p < 0.001 \)). No differences were observed across levels for C = 1 (\( \chi^2(1) = 1.93, p = 0.165 \)) and C = +1 (\( \chi^2(1) = 3.23, p = 0.072 \)).

Fig. 2 shows the ASSR magnitude in response to the gc-IR and its temporally reversed version, gc-IR\textsubscript{rev}. The upper panels represent the data obtained for the individual subjects at 50 dB SPL (left) and 70 dB SPL (right). As in Fig. 1, the SNR of the different recordings is represented using a gray scale, in dB, and the single open symbol indicates that the corresponding response was not significant in the condition gc-IR for subject MN. The lower panels in Fig. 2 show the ASSR magnitude relative to the mean of the two responses. At both stimulus levels, the response amplitude
obtained with gc-IR and gc-IRrev did not differ significantly from each other ($\chi^2(1) = 2.96$, $p = 0.085$ for 50 dB SPL; $\chi^2(1) = 0.46$, $p = 0.496$ for 70 dB SPL).

### 3.2. Simulation of frequency selectivity and cochlear dispersion

Fig. 3 shows the simulated BM responses to the Schroeder tone complexes. The solid black waveform represents the response of the filter tuned to the center frequency of the tone complex (1 kHz), referred to as the “on-frequency” response. The responses at the four other frequencies are shown as gray functions and represent “off-frequency” responses. The upper panel in Fig. 3 shows the simulations for the $\frac{C}{C_0} = 1$ condition. The middle and low panels show the corresponding results for $\frac{C}{C_0} = 0$ and $\frac{C}{C_0} = 1$, respectively. Above each of the three panels, the corresponding stimulus waveforms are shown for comparison.

The filtered stimuli in the case of $\frac{C}{C_0} = 1$ (upper panel) exhibit relatively flat envelopes in the individual frequency channels. For $\frac{C}{C_0} = 0$ (middle panel), the “response” envelopes are more peaky in the different filters and the response maxima are shifted in time across frequency. The temporally most compact response at the center frequency (1 kHz) can be seen for $\frac{C}{C_0} = 1$ (bottom panel). However, this effect is less pronounced, or absent, in the off-frequency channels. In particular, for this condition, there is a large phase shift of the individual channel responses across frequency.

Fig. 4 shows the simulated responses to gc-IR (left panel) and its temporally reversed version (gc-IRrev; right panel). The stimulus waveforms are indicated above the filtered model outputs. The differences between the responses to gc-IR and gc-IRrev are more subtle than in the case of the Schroeder-tone complexes since the stimulus bandwidth is much smaller than for the impulse response stimuli. Nevertheless, the same principal pattern can be observed in the results: First, the simulated response at the stimulus center frequency (2 kHz) is temporally more compact for gc-IRrev than for gc-IR. Second, the response peaks in the off-frequency channels provide stronger phase shifts across frequency in the case of the gc-IRrev stimulation than in the case of the gc-IR stimulation.

### 4. Discussion

ASSR with narrowband stimuli, like amplitude modulated tones, have been demonstrated to represent an objective estimate of frequency-specific sensitivity (e.g., Purcell et al., 2004; Herdman and Stapells, 2001). The present study tested, inspired by psychoacoustic masking data and auditory modeling studies, if ASSR to specific stimuli would be suitable for estimating effects of local (frequency-specific) cochlear dispersion in humans.

The hypothesis was that the Schroeder tone complex with positive phase curvature ($\frac{C}{C_0} = 1$) would compensate for phase delays within the transfer range of the BM filters at and around
the chosen stimulus center frequency (of, in this case, 1 kHz). It was anticipated that such processing would lead to a peaky internal representation after cochlear filtering at the center frequency, consistent with corresponding physiological near-field recordings in the cochlear nucleus of the chinchilla (Recio, 2001) and in line with the simulations of the filtered stimuli at the output of the BM (Fig. 3). The obtained ASSR data from the present study (Fig. 1) demonstrated, however, that the Schþ stimulus (C = 1) actually produced the smallest response amplitude whereas the temporally reversed stimulus (Sch−, C = −1) produced the largest response amplitude. Moreover, no differences were observed between the (normalized) ASSR magnitudes obtained with the Schþ stimulus (C = 1) at the two different sound pressure levels (left versus right panel in Fig. 1). In contrast, the psychophysical data from Oxenham and Dau (2001b) demonstrated a large effect of level for corresponding Schþ stimuli suggesting a level dependent (auditory filter) phase function. If the ASSR would be sensitive enough to reflect (on-frequency) dispersive effects, a smaller ASSR magnitude should be expected for the lower level than for the higher level, consistent with the hypothesis of a fast-acting compression mechanism that is more effective at the lower level (Carlyon and Datta, 1997). The results demonstrate that the chosen method and/or stimulation paradigm are not appropriate, or not sensitive enough, to represent frequency-specific cochlear phase effects.

The simple simulations with the linear BM model (Figs. 3 and 4) suggested that, even though dispersive effects in the on-frequency channel may be represented at a peripheral level, the across-frequency asynchrony of the BM excitation may dominate the “summed” response. It is important to note that a proper simulation of the ASSR would require a more realistic model of auditory peripheral and retro-cochlear processes as well as a transformation of the neural firing rates at the neural generator sites via convolution with the “unitary response” to the measured response at the electrodes in the far field (e.g. Melcher and Kiang, 1996; Dau, 2003; Ranne et al., 2012; Verhulst et al., 2015). Such more detailed modeling was not attempted here since the main motivation of the present study was to evaluate experimentally if the evoked potentials to the sequences of (bandlimited) chirps can represent properties of local cochlear dispersion.

Auditory evoked far-field responses represent synchronization potentials that are more pronounced the larger the number of simultaneously activated neurons. This is particularly evident for responses that are mainly generated at peripheral and brainstem levels, such as compound action potentials (CAPs), auditory brainstem responses (ABR) and auditory steady-state responses (ASSR). The role of synchrony of neural activity across the whole cochlear partition has been demonstrated in ABR and ASSR studies by using broadband rising chirps instead of clicks of similar bandwidth. If the stimulus compensates for the travel-time differences across frequency, as in the case of the rising chirp, the corresponding evoked potential exhibits a larger amplitude than in the case of the click and falling chirp stimulation (Dau et al.,

![Fig. 2. ASSR magnitude in response to the gc-IR at 2 kHz for stimulus levels of 50 dB SPL (panel A) and 70 dB SPL (panel B). The data from the individual subjects are shown in the upper panels. The gray scale represents the SNR of the recordings, in dB. The lower panels show the normalized response magnitude relative to the mean of the response for each subject.](image-url)
2000; Elberling and Don, 2008), despite the fact that the individual within-channel contributions are less compact in time in the case of the rising chirp compared to those of the other stimuli (e.g., Uppenkamp et al., 2001; Buss et al., 2002). In the case of the stimuli chosen in the present study to compensate for local dispersion, still the ASSR response is dominated by the synchronized activity across frequency which is more effective for the stimulus that exhibits smaller phase delays in the filtered representations across frequency (the Sch-stimulus) than for the Sch stimulus. This is the case for the 1800-Hz wide Schroeder tone complexes and is also still represented in the simulations with the narrowband IR stimuli (Fig. 4) even though the corresponding ASSRs (Fig. 2) with these stimuli did not show any significant differences.

It is possible that other stimulation paradigms, potentially with additional noise masking to suppress “off-frequency” contributions, are more effective than the ones considered here. Furthermore, higher-level, cortical, responses to tonal target tones in the presence of the different types of Schroeder tone complexes (as masks) have been demonstrated in a magneto-encephalography (MEG) study to be correlated with the corresponding behavioral masked thresholds of the tones (Rupp et al., 2002). The investigation provided in the present study demonstrates that the brainstem responses, as reflected in the ASSR, do not reflect frequency-specific cochlear dispersion despite the fact that ASSR are commonly used to estimate frequency specific hearing sensitivity.

5. Conclusion

ASSRs to Schroeder tone complexes that had earlier been used in behavioral masking studies to estimate auditory filter phase response were found in the present study to be dominated by “across-channel” synchrony and, thus, do not reflect local place-specific information. In the case of the more frequency-specific impulse response stimuli, no significant differences were found between the responses to the rising versus falling frequency sweeps. Thus, whereas ASSRs to narrowband stimuli have been considered as an objective indicator of frequency-specific hearing sensitivity, the paradigms tested in the present study do not seem to be sensitive enough to reflect local cochlear dispersion.

Acknowledgment

We would like to thank our colleagues from the Hearing Systems Group for valuable comments and stimulating discussions. We also thank the reviewers for their very helpful and constructive feedback on an earlier version of this paper. This work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (ChESs).

References


Carney, L.H., McDuffy, M.J., Shekhter, I., 1999. Frequency glides in the impulse