The effect of presentation level on spectrotemporal modulation detection

Sara Magits^{a,1}, Arturo Moncada-Torres^{a,1}, Lieselot Van Deun^{a,b},
Jan Wouters^a, Astrid van Wieringen^a, Tom Francart^{a,*}

^a*KU Leuven - University of Leuven, Dept. of Neurosciences, ExpORL, Herestraat 49,
Bus 721, 3000 Leuven, Belgium*

^b*Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals
Leuven, Leuven, Belgium*

Abstract

The understanding of speech in noise relies (at least partially) on spectrotemporal modulation sensitivity. This sensitivity can be measured by spectral ripple tests, which can be administered at different presentation levels. However, it is not known how presentation level affects spectrotemporal modulation thresholds. In this work, we present behavioral data for normal-hearing adults which show that at higher ripple densities (2 and 4 ripples/oct), increasing presentation level led to worse discrimination thresholds. Results of a computational model suggested that the higher thresholds could be explained by a worsening of the spectrotemporal representation in the auditory nerve due to broadening of cochlear filters and neural activity saturation. Our results demonstrate the importance of taking presentation level into account when administering spectrotemporal modulation detection tests.

*Corresponding author.

Email address: `tom.francart@kuleuven.be` (Tom Francart)

¹These authors declare equal contributions to this study and should be considered joint first authors.

Keywords: computational model, normal hearing, spectral ripple test, spectrotemporal sensitivity

1. Introduction

Complex acoustic signals such as speech are characterized by a combination of spectral and temporal modulations. Speech understanding relies (at least partially) on the ability to detect and discriminate these modulations. In other words, it relies on an individual’s spectrotemporal modulation sensitivity (Supin et al., 1997). This can be assessed by two categories of tests: *spectral ripple discrimination* tests and *spectral/spectrotemporal modulation detection* (SMD/STMD, respectively) tests. There are many varieties of these. However, in this paper we focus on an SMD/STMD test where participants are asked to discriminate between a modulated and unmodulated stimulus. The modulation detection threshold is usually defined as the minimal peak-to-valley ratio or modulation index at which the participant can discriminate between the two stimuli (e.g., Bernstein et al., 2013).

It has been shown that SMD/STMD thresholds are correlated with different measures of speech perception in quiet and in noise (Anderson et al., 2012; Mehraei et al., 2014; Davies-Venn et al., 2015; Croghan and Smith, 2018). Additionally, SMD/STMD thresholds can provide a non-linguistic measure of spectral/spectrotemporal sensitivity without the confounding factor of language knowledge that plays a role in standardized tests (e.g., speech audiometry, Gifford et al., 2014; Davies-Venn et al., 2015; Choi et al., 2016). This has motivated their use for a variety of purposes. For example, STMD paradigms have been used to explore perceptual learning mechanisms in the auditory system (Sabin et al., 2012). SMD/STMD tests have also used spectral/spectrotemporal resolution successfully as an outcome measure in different fields of audiological research: prediction of speech understanding in

noise of hearing-aid users (Bernstein et al., 2016), assessment of cochlear implant candidacy, parameter fitting, and new sound processing strategies (Langner et al., 2017; Choi et al., 2016; Croghan and Smith, 2018; Zheng et al., 2017), evaluation of bimodal hearing benefit (Zhang et al., 2013), and music perception (Choi et al., 2018).

Although these tests are used mostly in audiological research, to our knowledge no studies have evaluated how presentation level affects SMD/STMD thresholds. This is relevant because SMD/STMD thresholds might be negatively affected by the broadening of the auditory filters with increasing presentation level (Glasberg and Moore, 2000). Taking the effect of level into account is crucial when administering SMD/STMD tests in a research environment, in (potential) clinical practice, and even more in test situations where it cannot be controlled strictly (e.g., home-based computerized rehabilitation programs). Furthermore, we need to understand this effect to be able to make a fair comparison of behavioral SMD/STMD results obtained at different presentation levels within and across studies.

The goal of this work was to explore how presentation level affects SMD/STMD thresholds for young adult NH participants. Specifically, we focused on the STMD test, since spectrotemporally modulated (i.e., moving spectral ripple) stimuli have been suggested to provide a better representation of speech (Won et al., 2015) than stimuli measuring sensitivity to only spectral (i.e., rippled, Litvak et al., 2007; Saoji et al., 2009) or temporal modulation. Additionally, STMD tests prevent participants from having access to phase cues by using low rate temporal modulation (Bernstein et al., 2013). Furthermore, we used a biologically inspired model of peripheral processing up to the au-

51 ditory nerve (AN) to help us interpret the behavioral results, to study the
52 contribution of peripheral information to spectrotemporal sensitivity, and to
53 generate STMD threshold predictions.

54 **2. Behavioral Measurements**

55 *2.1. Materials & Methods*

56 *2.1.1. Participants*

57 Ten participants (1 male, 9 female, median age 23.5 years, age range
58 21–29 years) took part. They had audiometric thresholds ≤ 20 dB HL at
59 all octave frequencies from 125 to 8000 Hz. Written informed consent was
60 obtained. The study was approved by the Ethics Committee of the University
61 Hospitals Leuven (approval no. B322201731501).

62 *2.1.2. Equipment*

63 Measurements were performed in a double-walled sound-attenuating booth.
64 Stimuli were played from a computer via an RME Hammerfall DSP Multi-
65 face II sound card and presented to the participants through Sennheiser HDA
66 200 headphones using APEX 3 ([Francart et al., 2008](#)).

67 *2.1.3. Stimuli*

68 We used the spectrotemporally modulated stimuli described by [Kowalski](#)
69 [et al. \(1996\)](#) and [Chi et al. \(1999\)](#). These were 500-ms long (including 20-ms
70 onset and offset cosine ramps) and were generated with a sampling frequency
71 of 44100 Hz and 16-bit resolution using MATLAB (Mathworks, Natick, MA).

72 The spectral modulation was achieved as follows. The spectrum of the
73 ripple stimulus (the “carrier”) consisted of 4000 random-phase tones equally

74 spaced along the (logarithmic) frequency axis from 354 to 5656 Hz. The
 75 amplitudes of the individual components were adjusted to form a sinusoidally
 76 shaped spectrum around a flat base. The amplitude of the ripple was defined
 77 as the modulation depth m . The initial phase of the ripple Φ was defined
 78 relative to a sine wave starting at the low-frequency edge. Its value was set
 79 using 50 different selections of random phases between 0 and 2π to prevent
 80 participants from using phase differences as a cue. The ripple density was
 81 defined as Ω (with values of 0.5, 2, and 4 ripples/oct). The mathematical
 82 expression for the static ripple is given in Eq. 1

$$S(x) = 1 + m \sin(2\pi \Omega x + \Phi) \quad (1)$$

83 where x is the position on the logarithmic frequency axis (in octaves), which
 84 was defined as $x = \log_2(\frac{f}{f_0})$ with f being the component tone frequency and
 85 f_0 the low-frequency edge. Notice that when $m = 0$, the resulting profile is
 86 a flat spectrum.

87 The temporal modulation was achieved by moving the static ripple down-
 88 wards along the frequency axis at a constant velocity ω (defined as the num-
 89 ber of ripple per second passing the low-frequency edge of the spectrum).
 90 The value of ω was 4 Hz. The complete mathematical expression for the
 91 spectrotemporal modulated stimuli is given in Eq. 2, where t is time.

$$S(x, t) = 1 + m \sin(2\pi (\omega t + \Omega x) + \Phi) \quad (2)$$

92 In order to make our results comparable to those of previous studies, we
 93 report the modulation depth m as $20 \log_{10}(m)$ (i.e., in dB). The *reference*
 94 *stimulus* was unmodulated (i.e., $20 \log_{10}(m) = -\infty$ dB), whereas the modu-

95 lation depth of the *target stimulus* was varied adaptively (Sec. 2.1.4). Figure 1
96 shows spectrograms of the reference stimulus and two example target stimuli
97 ($20 \log_{10}(m) = -6$ dB and $20 \log_{10}(m) = 0$ dB).

98 2.1.4. Procedure

99 Initially, the stimuli were presented at levels of 65 and 86 dB SPL using all
100 three ripple densities (0.5, 2, and 4 ripples/oct). Then, stimuli were presented
101 at levels of 55, 65, 75, and 86 dB SPL with a ripple density of 4 ripples/oct.

Figure 1: (Color online) Spectrograms of the spectrotemporal modulated stimuli. Notice how the pattern along the frequency axis changes with increasing ripple density.

Throughout, stimuli were presented monaurally to the left ear. Level roving of 8 dB was used (i.e., random gain between -4 and 4 dB for each stimulus) to reduce the salience of level cues (Eddins and Bero, 2007).

A two-interval two-alternative forced-choice task was used. One of the intervals contained the unmodulated (i.e., reference) stimulus and the other interval contained the modulated (i.e., target) stimulus. The target was randomly presented in the first or second interval with equal probability. There was a 500-ms pause between intervals. Participants were seated in front of a computer screen. They were instructed to discriminate the target interval, which would correspond to the stimulus with a “rippled, vibrating sound”, from the reference interval, which would correspond to the stimulus with a “noisy sound”. They did so by clicking on the corresponding button on the screen (or by using the corresponding keys on the keyboard). Visual feedback was provided through a green (correct response) or red (incorrect response) highlight after each trial. Conditions were presented to each participant in a random order. In a given run, the ripple density was fixed. The modulation depth at threshold was estimated using a three-down one-up procedure tracking the 79.4% point on the psychometric function (Levitt, 1971). Each run started with a fully modulated target ($20 \log_{10}(m) = 0$ dB). The modulation depth was decreased by 6 dB after the first reversal, changed by 4 dB until two more reversals occurred, and changed by 2 dB for the last 6 reversals. A run was ended after 9 reversals. For each run, the mean value of $20 \log_{10} m$ at the last 6 reversals was calculated. Participants completed a test and retest run for every condition. If the thresholds for the two differed by more than 3 dB, a third run was completed. For each condition, the final threshold was

127 taken as the average of all runs.

128 2.2. Results

129 Statistical analysis was conducted using the R programming language
130 and statistical environment (R Core Team, 2017).

Figure 2: (Color online) STMD thresholds boxplot. The line in the middle of each box represents the median of the participants' thresholds for each condition. Symbols represent the average. The vertical edges of each box represent the 25th and 75th percentiles. The distance between them is the interquartile range (IQR). Error bars (i.e., whiskers) are drawn from the ends of the IQR to the furthest data point within 1.5 of the IQR. Crosses represent data points beyond that (i.e., outliers). Lower thresholds indicate better performance. * = $p < 0.05$, *** = $p < 0.001$.

131 Figure 2 shows a boxplot of the STMD thresholds together with the aver-
132 age across participants for stimuli at 65 and 86 dB SPL and 0.5, 2 and 4 rip-
133 ples/oct. A general linear model (GLM) showed that ripple density had a

134 significant effect on the STMD thresholds ($\chi^2(1) = 8.26, p < 0.001$) as
 135 did level ($\chi^2(1) = 11.76, p < 0.001$). There was a significant interaction
 136 of ripple density and level ($\chi^2(1) = 24.17, p < 0.001$). Tukey *post hoc*
 137 tests on the GLM revealed increased thresholds with increasing ripple den-
 138 sity at 86 dB SPL, between 0.5 ripples/oct and 4 ripples/oct ($z = 5.83,$
 139 $p < 0.001$, confidence interval (CI) [3.64, 8.02]) and between 2 ripples/oct
 140 and 4 ripples/oct ($z = 4.82, p < 0.001$, CI [2.63, 7.01]). In contrast,
 141 thresholds decreased with increasing ripple density at 65 dB SPL between
 142 0.5 ripples/oct and 2 ripples/oct ($z = -3.57, p < 0.001$, CI [-5.76, -1.38])
 143 and then increased between 2 ripples/oct and 4 ripples/oct ($z = 2.54,$
 144 $p = 0.012$, CI [0.35, 4.73]). The STMD thresholds were significantly lower
 145 at 65 dB SPL than at 86 dB SPL at 2 ripples/oct ($z = 3.12, p < 0.001$,
 146 CI [0.93, 5.31]) and at 4 ripples/oct ($z = 5.40, p < 0.001$, CI [3.21, 7.59]),
 147 but not at 0.5 ripples/oct ($z = -1.45, p = 0.40$, CI [-3.64, 0.73]).

148 Figure 3 shows a boxplot of the STMD thresholds for stimuli at 55, 65, 75, and
 149 86 dB SPL and 4 ripples/oct. STMD thresholds changed significantly with
 150 level (Friedman's ANOVA, $\chi^2_F(3) = 24.36, p < 0.001$). There was a large
 151 increase between 65 and 75 dB SPL. *Post hoc* Conover's tests with Holm
 152 correction for multiple comparisons revealed significant differences between
 153 55 and 75 dB SPL ($p < 0.001$), 55 and 86 dB SPL ($p < 0.001$), 65 and
 154 75 dB SPL ($p < 0.001$), 65 and 86 dB SPL ($p < 0.001$), and 75 dB SPL and
 155 86 dB SPL ($p < 0.001$). There was no significant difference between thresh-
 156 olds for the two lowest levels (55 and 65 dB SPL, $p > 0.05$).

Figure 3: (Color online) STMD thresholds boxplot with a ripple density of 4 ripples/oct. Otherwise as Figure 2.

157 3. Computational Model

158 We used a computational model with a physiologically inspired front end
 159 (i.e., model of the auditory periphery up to the AN) to help us interpret
 160 the behavioral results, to study the contribution of peripheral information
 161 to spectrotemporal sensitivity, and to obtain quantitative predictions of the
 162 behavioral thresholds. A block diagram of the model is shown in Fig. 4. We
 163 hypothesized that the model would reflect a worsening in the spectrotemporal
 164 representation in the AN with increasing level.

165 3.1. Stimuli

166 We included a wider range of levels (from 40 to 95 dB SPL in steps of
 167 5 dB) than used in the experiments. We used the same ripple densities

(0.5, 2, and 4 ripples/oct). We simulated responses to the reference stimulus
 $(20 \log_{10}(m) = -\infty \text{ dB})$ and target stimuli with a modulation depth of
 $20 \log_{10}(m) = -6 \text{ dB}$ and $20 \log_{10}(m) = 0 \text{ dB}$ (which correspond to 50 and
100% modulation, respectively, in a linear scale).

3.2. AN Model

The model proposed by Zilany et al. (2009, 2014) was used as a front
end. This model reproduces the responses of AN fibers to acoustic stimu-
lation. It has been validated with a wide range of physiological data. It

Figure 4: (Color online) Block diagram of the computational model used to interpret the behavioral data and to study the contribution of peripheral information to spectrotemporal sensitivity.

176 is comprised of different modules (each simulating a specific function of the
177 auditory periphery).

178 First, the stimulus is passed through a filter simulating the middle ear
179 frequency response. The output is fed to a signal path and a control path.
180 The signal path mimics the behavior of the outer-hair-cell- (OHC-) controlled
181 filtering of the basilar membrane in the cochlea and the transduction of the
182 inner-hair-cells (IHCs) by a series of non-linear and low-pass filters. The con-
183 trol path mimics the function of the OHCs in controlling basilar membrane
184 filtering. The control path output feeds back into itself and into the signal
185 path. The output of the IHCs is fed to the IHC-AN synapse module with
186 two power-law adaptation paths, which simulate slow and fast adaptation.

187 For each stimulus, the AN model generated a so-called early stage neu-
188 rogram (ESN). An ESN is a time-frequency representation of a signal which
189 encodes temporal modulations caused by the interaction of spectral compo-
190 nents in each band ([Elhilali et al., 2003](#)). It shows the response of neurons
191 tuned to different characteristic frequencies (CFs) through time. We used
192 512 CFs logarithmically spaced from 250 to 8000 Hz. For each CF, we simu-
193 lated the average response of 50 AN fibers with different spontaneous rates:
194 high (100 spikes/s), medium (5 spikes/s), and low (0.1 spikes/s), with propor-
195 tions of 0.6, 0.2, and 0.2, respectively, which correspond to the distribution
196 observed in mammals ([Liberman, 1978](#); [Zilany and Bruce, 2007](#)). We grouped
197 the neural activity into time bins of 8 ms, which is close to the equivalent
198 rectangular duration of the temporal window of the auditory system ([Moore
199 et al., 1988](#); [Oxenham and Moore, 1994](#)). Afterwards, we smoothed the re-
200 sponse by convolving it with a 2-sample long rectangular window with 50%

201 overlap. Figure 5 shows example ESNs of reference and target stimuli for
202 different ripple densities.

Figure 5: (Color online) ESNs of spectrotemporally modulated stimuli.

203 3.3. Neurogram activity

204 We quantified the increase of neural activity by computing the mean and
205 standard deviation of the neurograms across different levels. Figure 6 shows
206 plots of the ESN activity for the reference stimulus ($20 \log_{10}(m) = -\infty$ dB)
207 and a fully-modulated target stimulus ($20 \log_{10}(m) = 0$ dB). In all cases,

208 increasing the level increased the neural activity and the slope of the curves
 209 decreased at high levels. These trends were consistent across all three ripple
 210 densities.

Figure 6: (Color online) Neurogram activity. The solid lines and the shaded areas correspond to the mean and standard deviation, respectively, of each neurogram. This is a measure of the amount of activity at the AN level. A large increase in activity could lead to saturation and, therefore, to a poorer spectrotemporal representation, yielding higher thresholds (Sec. 4). The dashed lines represent the levels at which behavioral measurements were obtained.

211 3.4. Neurogram frequency profiling

212 We defined a *frequency profile* of a neurogram as a slice across its CFs at
 213 a given point in time. If we think of a neurogram as an image, a frequency
 214 profile would correspond to all the row values of a specific column.

215 Consider the ESNs in Fig. 5 for the ripple density of 0.5 ripples/oct.
 216 The top ESN ($20 \log_{10}(m) = -\infty$ dB) shows a uniform, indistinct pattern.
 217 A frequency profile at any point in time would show a roughly flat curve.
 218 In contrast, the bottom ESN ($20 \log_{10}(m) = 0$ dB) shows a clear pattern,
 219 reflecting the spectrotemporal characteristics of the stimulus. A frequency

220 profile at any point in time would show distinct crests and troughs. Figure 7
221 shows frequency profiles for different ripple densities at different levels.

222 3.5. Dispersion

223 One measure of the information available at the AN level for detection
224 of modulation is the dispersion of the frequency profiles (i.e., columns) of
225 the ESNs across time. The dispersion is a measure of the amplitude of
226 the frequency profile curves. It measures the amount of variation in am-
227 plitude across the frequency range. We quantified this dispersion using the
228 interquartile range (IQR), as shown in Eq. 3. We also computed a measure
229 of the dispersion variability across all the frequency profiles of a given neu-
230 rogram, as shown in Eq. 4. In both cases, ESN_j is the frequency profile at
231 the j -th point in time.

$$ESN_{disp} = \text{Median}(\text{IQR}(ESN_j)) \quad (3)$$

Figure 7: (Color online) Frequency profiles of the ESNs at $t = 250$ ms (total duration of the stimulus was 500 ms).

$$\text{ESN}_{\text{dispVar}} = \text{IQR}(\text{IQR}(\text{ESN}_j)) \quad (4)$$

Figure 8 shows plots of ESN dispersion. Deeper modulations (closer to $20 \log_{10}(m) = 0$ dB) led to larger dispersions for lower ripple densities (0.5 and 2 ripples/oct). In all cases, increasing the level reduced the dispersion. This trend was consistent across all three ripple densities.

Figure 8: (Color online) Plots of ESN dispersion. The solid lines and the shaded areas correspond to the ESN dispersion and dispersion variability, respectively, for frequency profiles across all time points for each neurogram. The ESN dispersion is a measure of the amount of information for modulation detection available at the AN level (larger dispersion allows for higher detectability, Sec. 4). The dashed lines represent the levels at which behavioral measurements were obtained.

3.6. Regression

Model results were compared with the behavioral data using a regression model. Since the results of experiment 1 showed that the effect of presentation level was largest at 4 ripples/oct, we focused on the behavioral data for experiment 2.

241 We calculated the difference in dispersion between a fully-modulated tar-
 242 get stimulus ($20 \log_{10}(m) = 0$ dB) and the non-modulated reference as a
 243 predictor for an exponential regression model as described by Eq. 5:

$$\text{Behav. thresh.}(\text{ESN}_{\text{disp}}, \text{ESN}_{\text{disp ref}}) = a e^{b(\text{ESN}_{\text{disp}} - \text{ESN}_{\text{disp ref}})} \quad (5)$$

244 with parameters a and b . It yielded an (adjusted) R^2 value of 0.98 and a
 245 root mean squared error (RMSE) of 0.25 dB. Figure 9 shows plots of the
 246 behavioral data versus the model metric as well as the regression model.
 247 We used the generated model to predict the behavioral thresholds for the
 248 different levels. Figure 10 shows the model's predictions as well as the mean
 249 of the behavioral data (as a reference). The model predictions show that the
 250 lowest (best) STMD threshold is around $20 \log_{10}(m) = -13.5$ dB for the
 251 modelled experiment.

Figure 9: (Color online) Exponential regression model of the behavioral thresholds of experiment 2 (ripple density of 4 ripples/oct).

Figure 10: (Color online) Model predictions (squares) of the behavioral thresholds (circles) across different levels (ripple density of 4 ripples/oct).

4. Discussion

Higher levels led to increased STMD thresholds. Moreover, increasing ripple density affected the STMD thresholds differently depending on the level. At 65 dB SPL, STMD thresholds were lowest at 2 ripples/oct. In other studies a similar trend was found. [Anderson et al. \(2012\)](#) found lowest thresholds at 3 ripples/oct, followed by increasing thresholds with increasing ripple density (up to 64 ripples/oct). The participants of [Eddins and Bero \(2007\)](#) performed best either at 2 or 3 ripples/oct. [Davies-Venn et al. \(2015\)](#) found a significant improvement in thresholds from 0.5 to 1 and from 1 to 2 ripples/oct. Other studies ([Bernstein and Green, 1987](#); [Leek and Summers, 1996](#)) have found similar trends. The most common explanation is that there are two regions in which different cues are used. For low ripple densities (≤ 3 ripples/oct), the ripples are detected using a spectral-contrast mechanism, while for higher ripple densities (> 3 ripples/oct), the spectral cues become weaker and the interaction between the close peaks in the rippled

noise provides usable temporal cues ([Davies-Venn et al., 2015](#)). However, further studies are needed to confirm this. At 86 dB SPL, STMD thresholds increased with increasing ripple density, similar to what [Bernstein et al. \(2013\)](#) found. The effect of presentation level was largest at 4 ripples/oct, where low presentation levels (55 and 65 dB SPL) yielded significantly lower (better) STMD thresholds than high presentation levels (75 and 86 dB SPL).

Understanding the effect of level on STMD thresholds for NH listeners is the first step to understanding it in HI listeners. Although it is very likely that level also affects STMD thresholds of HI listeners, our results cannot be translated directly to the HI population for several reasons. Firstly, increasing the intensity affects neural saturation for NH and HI listeners differently. This can also affect perception differently due to the abnormal loudness-growth curve (i.e., non-linear loudness shift, [Edwards et al., 1998](#); [Hellman, 1999](#)) of HI listeners. Additionally, the auditory filters of HI listeners are abnormally broad (resulting in spectral smearing of the internal representation of the stimulus, [Moore, 2007](#)) and change less with level compared to NH listeners. Furthermore, the large heterogeneity of the HI population ([Lopez-Poveda and Johannesen, 2012](#)) would very likely play a role. Therefore, we hypothesize that STMD thresholds of HI listeners will also be affected by level and will be worse than those of NH listeners. However, testing this requires further behavioral measurements and modelling. This would be a crucial step for further understanding the differences in STMD thresholds between NH and HI participants. Our results show that attributing them to differences in spectrotemporal sensitivity would be only partially true, since level also plays an important role.

We used a computational model with a physiologically inspired front end to explain the behavioral results (Fig. 5). We found that the observed effects of level on the behavioral data could be explained by a worsening of the spectrotemporal representation in the AN due to broadening of the cochlear filters. Furthermore, higher levels led to more neural saturation “filling in the dips” of the neurograms. This can be seen in the increase of the neural activity (Fig. 6) and the flattening of the frequency profiles (Fig. 7). Frequency profiles at lower levels reflected the changes of the spectral information across time, while frequency profiles at higher levels lost the representation of this information (Fig. 8). All these factors diminish the coding of the spectrotemporal pattern of the modulated stimuli in the AN with increasing level, making it harder to discriminate.

The regression analysis (Fig. 9) suggested that information in the auditory periphery is able to account for a large proportion of the variance in the behavioral data, supporting its value for predicting spectrotemporal modulation thresholds (Fig. 10).

Similar results could have been obtained with a more simple model (e.g., an excitation pattern model, Moore and Glasberg, 1987). However, the use of frameworks based on the biology of the auditory system has a few advantages. For instance, they incorporate physiological information inherently. This allows a more direct, transparent understanding of the auditory mechanisms at different stages of the auditory pathway (the periphery in this case), since it gives insight into the representation of the stimuli at each of these steps. Additionally, the Zilany et al. (2009, 2014) AN model incorporates the effects of sensorineural hearing loss due to damage to the IHCs

317 and OHCs (something that would not be straightforward to do using a non-
318 physiological approach). Now that presented framework has been validated
319 for the NH case, this would be of special interest, since it could allow studying
320 the effect of level on spectrotemporal modulation detection by HI listeners
321 using a similar framework to the one described here.

322 Furthermore, alternative back ends could have been used in the proposed
323 model. For example, the ratio between the dispersion of the reference and
324 the target stimulus (instead of the difference) could have been used as the
325 predictor for the regression. Additionally, a different approach could have
326 been used to predict the behavioral threshold. For instance, the difference
327 in dispersion (or the quotient) between the reference and the target stimulus
328 required for threshold could be computed. Afterwards, the modulation depth
329 required to achieve this difference metric could be calculated iteratively, with
330 the final value being the predicted behavioral threshold. This approach would
331 eliminate the need for the regression model in Eq. 5.

332 The effect of presentation level has a number of implications for the use of
333 STMD tests in experimental and clinical environments. When administering
334 STMD tests at different levels, the observed differences in STMD thresholds
335 should (at least partially) be attributed to the effect of level, making it
336 more complex to interpret the contribution of spectrotemporal sensitivity
337 only. For NH participants it is recommended to use a fixed presentation level
338 to allow for direct comparison between their STMD thresholds. However,
339 it is unclear how level affects STMD thresholds in HI listeners. Therefore
340 recommendations for STMD tests in HI participants cannot be made based
341 on our data. Future work will be focused on investigating level effects for

different types of spectral and spectrotemporal ripple tests, as well as for HI listeners.

5. Conclusions

STMD thresholds were higher (worse) at high than at low presentation levels, with larger differences in thresholds at 4 ripples/oct than at 2 ripples/oct. The computational model with a physiologically inspired front end could account for the behavioral results, showing that information at the peripheral level is sufficient to predict the behavioral thresholds. STMD thresholds obtained at different presentation levels are affected not only by differences in spectrotemporal modulation, but also at least partly by level. Therefore, this effect needs to be considered when administering STMD tests (both in clinical practice and in experimental research) and when comparing STMD thresholds within and across studies.

Acknowledgements

This work was supported by a TBM-FWO grant from the Research Foundation-Flanders [grant number T002216N] and an ERC Starting Grant (to Tom Francart) from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme [grant agreement no. 637424]. In addition, this research was jointly funded by Cochlear Ltd. and Flanders Innovation & Entrepreneurship (formerly IWT), project 50432. We would like to thank Benjamin Dieudonné for his technical help generating the stimuli. We would also like to thank the editor Dr. Brian C. J. Moore and the three anonymous reviewers for their helpful comments and feedback.

365 Finally, our most sincere thanks to our participants for their time and dedi-
366 cated effort.

367 Anderson, E. S., Oxenham, A. J., Nelson, P. B., Nelson, D. A., 2012. Assess-
368 ing the role of spectral and intensity cues in spectral ripple detection and
369 discrimination in cochlear-implant users. *The Journal of the Acoustical*
370 *Society of America* 132 (6), 3925–34.

371 Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M.,
372 Leek, M. R., 2013. Spectrotemporal modulation sensitivity as a predic-
373 tor of speech intelligibility for hearing-impaired listeners. *Journal of the*
374 *American Academy of Audiology* 24 (4), 293–306.

375 Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnerberg,
376 J., Lunner, T., nov 2016. Spectrotemporal Modulation Sensitivity as a
377 Predictor of Speech-Reception Performance in Noise With Hearing Aids.
378 *Trends in hearing* 20, 1–17.

379 Bernstein, L. R., Green, D. M., 1987. The profile-analysis bandwidth. *The*
380 *Journal of the Acoustical Society of America* 81 (6), 1888–1895.

381 Chi, T., Gao, Y., Guyton, M. C., Ru, P., Shamma, S., 1999. Spectro-
382 temporal modulation transfer functions and speech intelligibility. *Journal*
383 *of the Acoustical Society of America* 106 (5), 2719–2732.

384 Choi, J. E., Hong, S. H., Won, J. H., Park, H.-S., Cho, Y. S. Y.-S., Chung,
385 W.-H., Cho, Y. S. Y.-S., Moon, I. J., 2016. Evaluation of Cochlear Implant
386 Candidates using a Non-linguistic Spectrotemporal Modulation Detection
387 Test. *Scientific reports* 6, 35235.

- 388 Choi, J. E., Won, J. H., Kim, C. H., Cho, Y. S., Hong, S. H., Moon, I. J.,
389 2018. Relationship between spectrotemporal modulation detection and mu-
390 sic perception in normal-hearing, hearing-impaired, and cochlear implant
391 listeners. *Scientific Reports* 8 (1), 1–11.
- 392 Croghan, N. B. H., Smith, Z. M., 2018. Speech Understanding With Vari-
393 ous Maskers in Cochlear-Implant and Simulated Cochlear-Implant Hear-
394 ing: Effects of Spectral Resolution and Implications for Masking Release.
395 *Trends in Hearing* 22 (July), 233121651878727.
- 396 Davies-Venn, E., Nelson, P., Souza, P., 2015. Comparing auditory filter band-
397 widths, spectral ripple modulation detection, spectral ripple discrimina-
398 tion, and speech recognition: Normal and impaired hearing. *The Journal*
399 *of the Acoustical Society of America* 138 (1), 492–503.
- 400 Eddins, D. A., Bero, E. M., 2007. Spectral modulation detection as a function
401 of modulation frequency, carrier bandwidth, and carrier frequency region.
402 *The Journal of the Acoustical Society of America* 121 (1), 363–372.
- 403 Edwards, B. B. W., Struck, C. J., Dharan, P., Hou, Z., 1998. New Digital
404 Processor for Hearing Loss Compensation based on the Auditory System.
405 *The Hearing Journal* 51 (8), 49–52.
- 406 Elhilali, M., Chi, T., Shamma, S. A., 2003. A spectro-temporal modulation
407 index (STMI) for assessment of speech intelligibility. *Speech Communica-*
408 *tion* 41 (2-3), 331–348.
- 409 Francart, T., van Wieringen, A., Wouters, J., 2008. APEX 3: a multi-purpose

410 test platform for auditory psychophysical experiments. *Journal of Neuro-*
411 *science Methods* 172 (2), 283–293.

412 Gifford, R. H., Hedley-Williams, A., Spahr, A. J., 2014. Clinical assessment of
413 spectral modulation detection for adult cochlear implant recipients: A non-
414 language based measure of performance outcomes. *International Journal*
415 *of Audiology* 53 (3), 159–164.

416 Glasberg, B. R., Moore, B. C. J., 2000. Frequency selectivity as a function
417 of level and frequency measured with uniformly exciting notched noise.
418 *Journal of the Acoustical Society of America* 108 (5), 2318–2328.

419 Hellman, R. P., 1999. Cross-modality matching: A tool for measuring loud-
420 ness in sensorineural impairment. *Ear and hearing* 20 (3), 193–213.

421 Kowalski, N., Depireux, D. A., Shamma, S. A., 1996. Analysis of dynamic
422 spectra in ferret primary auditory cortex. I. Characteristics of single-unit
423 responses to moving ripple spectra. *Journal of Neurophysiology* 76 (5),
424 3503–3523.

425 Langner, F., Saoji, A. A., Büchner, A., Nogueira, W., 2017. Adding simul-
426 taneous stimulating channels to reduce power consumption in cochlear
427 implants. *Hearing Research* 345, 96–107.

428 Leek, M. R., Summers, V., 1996. Reduced frequency selectivity and the
429 preservation of spectral contrast in noise. *The Journal of the Acoustical*
430 *Society of America* 100 (3), 1796–1806.

431 Levitt, H., feb 1971. Transformed up-down methods in psychoacoustics. *The*
432 *Journal of the Acoustical society of America* 49 (2), 467–477.

- 433 Liberman, M. C., 1978. Auditory-nerve response from cats raised in a low-
434 noise chamber. The Journal of the Acoustical Society of America 63 (2),
435 442–455.
- 436 Litvak, L. M., Spahr, A. J., Saoji, A. A., Fridman, G. Y., 2007. Relationship
437 between perception of spectral ripple and speech recognition in cochlear
438 implant and vocoder listeners. The Journal of the Acoustical Society of
439 America 122 (2), 982–991.
- 440 Lopez-Poveda, E. A., Johannesen, P. T., 2012. Behavioral estimates of the
441 contribution of inner and outer hair cell dysfunction to individualized au-
442 diometric loss. Journal of the Association for Research in Otolaryngology
443 13 (4), 485–504.
- 444 Mehraei, G., Gallun, F. J., Leek, M. R., Bernstein, J. G. W., 2014. Spec-
445 trotemporal modulation sensitivity for hearing-impaired listeners: Depen-
446 dence on carrier center frequency and the relationship to speech intelligi-
447 bility. The Journal of the Acoustical Society of America 136 (1), 301–316.
- 448 Moore, B. C., Glasberg, B. R., Plack, C., Biswas, A., 1988. The shape of the
449 ears temporal window. The Journal of the Acoustical Society of America
450 83 (3), 1102–1116.
- 451 Moore, B. C. J., 2007. Cochlear Hearing Loss: Physiological, Psychological,
452 and Technical Issues. Wiley Series in Human Communication Science.
- 453 Moore, B. C. J., Glasberg, B. R., 1987. Formulae describing frequency se-
454 lectivity as a function of frequency and level, and their use in calculating
455 excitation patterns. Hearing research 28 (2-3), 209–225.

- 456 Oxenham, A. J., Moore, B. C., 1994. Modeling the additivity of nonsimulta-
457 neous masking. *Hearing research* 80 (1), 105–118.
- 458 R Core Team, 2017. R: A Language and Environment for Statistical Com-
459 puting. R Foundation for Statistical Computing, Vienna, Austria.
460 URL <https://www.r-project.org/>
- 461 Sabin, A. T., Eddins, D. a., Wright, B. a., 2012. Perceptual learning evi-
462 dence for tuning to spectrotemporal modulation in the human auditory
463 system. *The Journal of neuroscience : the official journal of the Society for*
464 *Neuroscience* 32 (19), 6542–9.
- 465 Saoji, A. A., Litvak, L., Spahr, A. J., Eddins, D. A., 2009. Spectral modula-
466 tion detection and vowel and consonant identifications in cochlear implant
467 listeners. *Journal of the Acoustical Society of America* 126 (3), 955–958.
- 468 Supin, A. Y., Popov, V. V., Milekhina, O. N., Tarakanov, M. B., 1997.
469 Frequency-temporal resolution of hearing measured by rippled noise. *Hear-*
470 *ing Research* 108 (1-2), 17–27.
- 471 Won, J. H., Moon, I. J., Jin, S., Park, H., Woo, J., Cho, Y. S., Chung, W. H.,
472 Hong, S. H., 2015. Spectrotemporal modulation detection and speech per-
473 ception by cochlear implant users. *PLoS ONE* 10 (10), 1–24.
- 474 Zhang, T., Spahr, A. J., Dorman, M. F., Saoji, A., 2013. Relationship Be-
475 tween Auditory Function of Nonimplanted Ears and Bimodal Benefit. *Ear*
476 *and hearing* 34 (2), 133–41.
- 477 Zheng, Y., Escabí, M., Litovsky, R. Y., 2017. Spectro-temporal cues enhance

- 478 modulation sensitivity in cochlear implant users. *Hearing Research* 351,
479 45–54.
- 480 Zilany, M. S. A., Bruce, I. C., 2007. Predictions of speech intelligibility with
481 a model of the normal and impaired auditory-periphery. In: *Neural Engi-*
482 *neering, 2007. CNE'07. 3rd International IEEE/EMBS Conference. IEEE,*
483 *Kohala Coast, HI, USA, pp. 481–485.*
- 484 Zilany, M. S. A., Bruce, I. C., Carney, L. H., 2014. Updated parameters and
485 expanded simulation options for a model of the auditory periphery. *The*
486 *Journal of the Acoustical Society of America* 135 (1), 283–286.
- 487 Zilany, M. S. A., Bruce, I. C., Nelson, P. C., Carney, L. H., 2009. A phe-
488 nomenological model of the synapse between the inner hair cell and audi-
489 tory nerve: Long-term adaptation with power-law dynamics. *The Journal*
490 *of the Acoustical Society of America* 126 (5), 2390–2412.