inter.noise
HAMBURG 2016

Modeling spectro-temporal modulation perception in normal-hearing listeners

Raul H. SANCHEZ[1]; Torsten DAU[1]

[1] Hearing Systems group, Department of Electrical Engineering,

Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

ABSTRACT

The ability of human listeners to detect and discriminate spectro-temporal ripples in sound has been shown to be correlated with speech intelligibility performance in several conditions. Thus, if a model would be able to account for the spectro-temporal processing limits in the auditory system, such a framework could be used to analyze the auditory processes contributing to and limiting speech intelligibility. Here, a model is presented that combines the concepts of the power spectrum model of masking (PSM; Patterson and Moore, 1986) with those of the speech based envelope power spectral model of masking (EPSM; Jørgensen and Dau, 2011). Effects of masking and changes in the signal-to-noise ratio in both domains are considered in the decision device of the model. The model was evaluated in experimental conditions of temporal, spectral and combined spectro-temporal modulation detection and discrimination using identical stimuli as input to the model as to the human listeners. The predictions were compared to the measured data obtained with 15 normal-hearing listeners. The model could account for the mean data in most of the considered conditions and might provide a valuable framework for investigating effects of hearing impairment both on spectro-temporal perception as well as speech intelligibility.

Keywords: Spectro-temporal modulation, auditory modeling. Number(s): 76.9, 78, 79.9

1. INTRODUCTION

Speech signals are quite dynamic in that they exhibit spectral and temporal modulations. The ability of human listeners to detect and discriminate these spectro-temporal ripples in sound has been shown to be correlated with speech intelligibility performance in several conditions (1–3). Speech prediction models based on spectro-temporal properties of speech provided accurate results (4–6), reproducing normal-hearing listeners data from speech-in-noise tests. Recently, Bernstein et al. (1) and Mehraei et al. (7) showed significant differences between normal and hearing impaired listeners in spectro-temporal modulation (STM) detection and its relation to speech intelligibility in noise. Thus, further investigation in terms of the limitations of STM perception could be interesting for audiological applications. Furthermore, if a model would be able to account for the spectro-temporal processing limits in the auditory system, such a framework could be used to analyze the auditory processes contributing to and limiting speech intelligibility.

The sensitivity to modulations has been studied in normal-hearing listeners (NH) using broadband noise, yielding temporal (T-MTFs), spectral (S-MTFs) and spectro-temporal modulation transfer functions (ST-MTFs) (8–10). T-MTFs have been characterized by a low-pass behavior where at low modulation frequencies (f_m), the detection threshold remains fairly constant and increases with a cutoff frequency of f_m = 64 Hz (8). In contrast, S-MTFs showed a band-pass characteristic, with a minimum located at specific spectral densities (number of spectral ripples per octave) that occurs at 2 to 4 cycles per octave, which means that the sensitivy is higher at these spectral densities. In the case of the spectro-temporal modulations, NH listeners were more sensitive at the same spectral densities as observed in the S-MTFs. Thus, Chi et al.(10) argued that ST-MTFs are the product of temporal and spectral detection, so they are separable. However, it seems that the sensitivity to STM decreases more rapidly than for spectral modulations when increasing the spectral density (10).

[1] tdau@elektro.dtu.dk

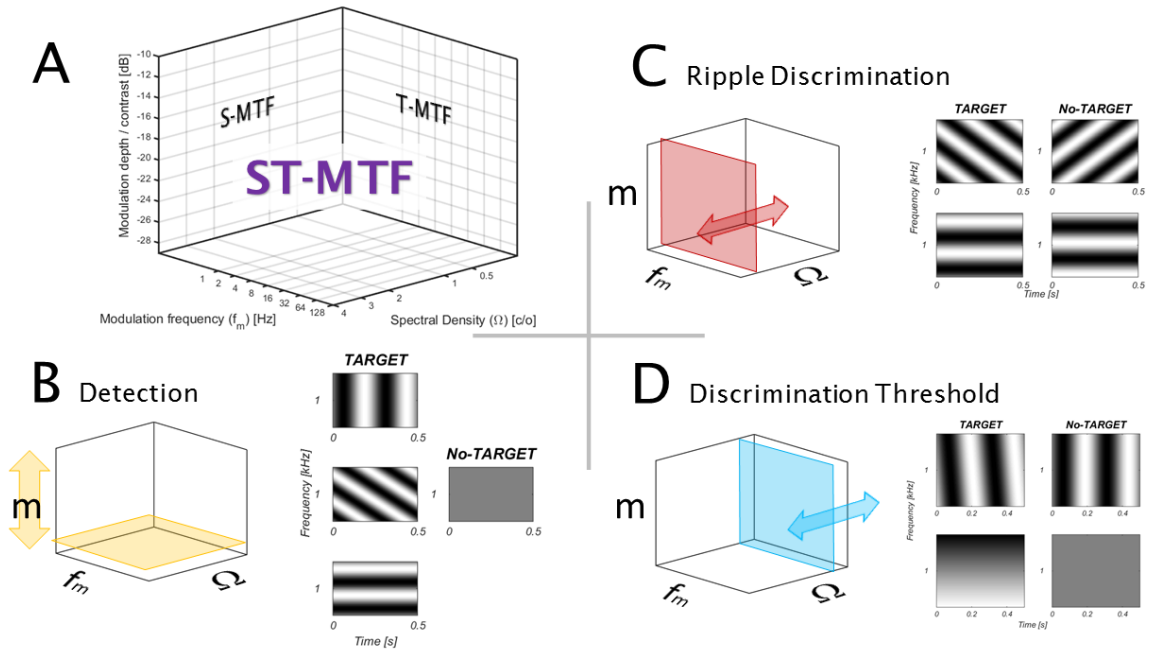

Figure 1 – Overview of the present study. A) Spectral, temporal and spectro-temporal modulation transfer functions. T-MTF corresponds to a ST-MTF where Ω = 0c/o and S-MTF corresponds to a ST-MTF where f_m = 0 Hz. B,C and D) The colored planes depict the experiments proposed here. Yellow plane shows the detection task (B) where the target is a modulated noise. Red plane shows the ripple discrimination experiments where two fully modulated with different patterns have to be discriminated (C). Blue plane the discrimination threshold where the task consist in discriminate between two stimuli when spectral density is added (D). As a result, the ST modulation perceptual limitations can be bounded in the three dimensions.

While S-MTF and ST-MTF using 1 octave band noise carriers showed similar trends as in the case of broadband noise (7,11), Dau et al. (12) observed that the spectral density of the inherit fluctuations of the carrier (i.e. its bandwidth) yielded different T-MTF patterns for narrow band noise. Specifically, when noise was limited to a single critical band, the temporal modulation detection could be simply explained by the difference between the modulation and the envelope power of the carrier, which led to the idea of the envelope power spectrum of masking (EPSM)(13). Later, Jørgensen and Dau (14) applied this idea in a speech prediction model which makes use of the signal-to-noise ratio of the envelope (SNRenv) as a metric of the speech intelligibility. The model consists of a peripheral filter-bank, an envelope extraction stage, and a modulation filter-bank that analyses the envelope of the output of each auditory filter. Although the results of this speech based model showed a good agreement with the human data, this approach has not been used to reproduce S-MTF or ST-MTF yet.

Chi et al. (10) proposed a model that analyses the auditory spectrogram -spectrogram based on a biological inspired auditory processing- by a cortical bank of modulation filters which were tuned to different combinations of modulation rates and spectral densities. This stage is biologically inspired by the responses of the auditory primary cortex, which exhibit selectivity to spectro-temporal modulations, so-called spectro-temporal receptive fields (STRF). This approach has also a speech-based extension, the spectro temporal modulation index (STMI), which was able to reproduce normal-hearing listeners data in different acoustic conditions (5). Recently, Bernstein et al. (4) attempted to reproduce STM detection using a similar approach. The individual data of NH and HI listeners were used to tune the model to a certain STM detection condition. This model successfully predicted the speech reception thresholds of both groups. However, the model failed in reproducing the other STM conditions at higher rate/density combinations. Although the STRF may be needed to explain the segregation of sounds in complex scenarios, here, the use of models based on the classical theories of power spectral model of masking (15) and its equivalent in the envelope spectrum domain (EPSM)

(16) will be investigated. The objective is to clarify to what extent, a basic auditory signal processing can account for the spectral, temporal and spectro-temporal combinations.

As mentioned above, the ability of perceiving speech in noise has been also connected to the ability to discriminate spectral ripples. The spectral ripple discrimination (SRD) experiment carried by Henry et al. (2) showed that HI listeners had a reduced spectral ripple discrimination as happened in listeners with cochlear implants. The task consisted of detecting the interval that contains a spectral ripple, modulated with the same modulation depth, but with the peaks and valleys of the spectrum reversed. However, the mechanisms involved in detection and discrimination tasks have been argued to be different for spectral ripples (3,17). In part, this is because studies involving these stimuli are often carried out using broadband stimuli. Despite spectral ripple discrimination being a time efficient and nonlinguistic task connected to the speech intelligibility (17), there are not systematic studies that could show the human limitations to perceive this stimuli in bandlimited stimuli. Therefore, it would be interesting to clarify the relationship between modulation detection and discrimination and the contribution of temporal and spectral cues involved in the modulation sensitivity of ripples and the discrimination between TM and STM, was also studied here.

The present study attempts to clarify the perceptual limitations observed in NH listeners in terms of the detection of modulations, the minimum differences in type of modulations (discrimination thresholds) or the pattern of the modulation (ripple discrimination) using 1 octave band carriers at 1 and 4 kHz (see Figure 1). Moreover, a model based on classical power spectrum models was used to partially explain modulation perception in several tasks. The purpose of this modeling approach is to examine the limitations of an "efficient" model based on psychoacoustic experiments and only fitted by only one parameter. The main hypothesis addressed here is that the combination of peripheral and modulation filters is already able to explain the majority of the conditions because their implementation is based in temporal resolution and frequency selectivity.

2. Basic auditory-filter model

The model acts as an ideal observer, which performs the experiments in the same way as the participants of this study. All the psychoacoustical tasks were carried out using a 3-interval forced-choice (3IFC) adaptive paradigm and the listeners were asked to identify the interval that contained the sound that was perceived to be more different than the other two. In the present model, the signal of each interval was processed by an auditory processing stage followed by a decision device that quantifies the differences among the intervals using the interval-to-interval ratio (I2IR), which was based on the combination of the signal-to-noise ratio envelope (SNR_{env}) (14) and the optimal detector described in (18).

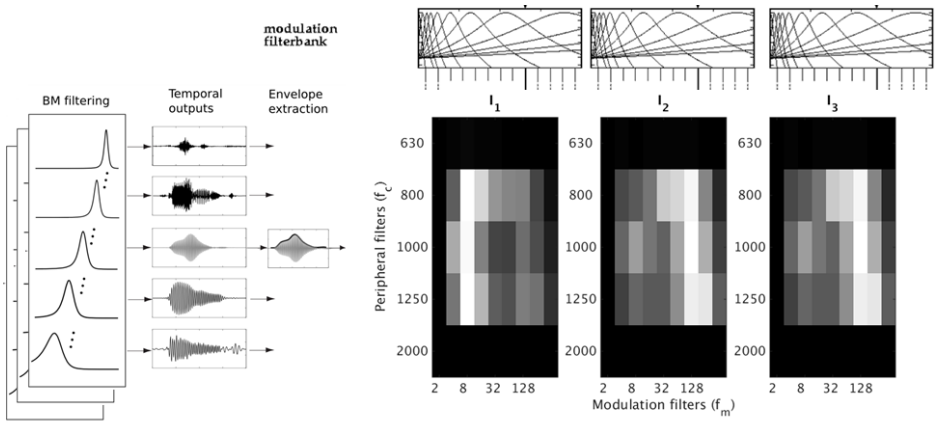

Figure 2 – Block diagram of the model. Signals of the three intervals are processed (auditory filter-bank, envelope extraction and modulation filter-bank. As a result a GxM matrix represents the internal representation of each of the intervals.

2.1 Front-end: Auditory signal processing

Figure 2 illustrates the stages of the front-end of the auditory model. The signal presented in each of the intervals is first processed by an auditory filter-bank (19) that divides the input in G spectral channels (x_g). Subsequently the envelope is subtracted ($xenv_g$) and for each auditory channel, this is analyzed by a modulation filter-bank that filters the envelope spectrum by using M bandpass filters (20). The output of the front-end is a three-dimensional time-varying signal ($Xenv_{g,m}$) that will be further analyzed in the back-end.

The auditory filter-bank used here is a gammatone filter that simulates the basilar membrane bandpass-filter characteristics. The filter-bank consists of 24 filters equally spaced by means of the equivalent rectangular bandwidth (ERB) scale (21). Only the filters that are considered "audible" (less than 20 dB below the level in the band that contains the highest power) will be used in further stages. For each x_g, the signal is half-wave rectified and down-sampled with new sampling frequency of 3 kHz. This processing filters the rectified signal at 1.5 kHz, which preserves the temporal fine structure only in the low frequencies range while reducing the computational cost in the later stages. Each envelope is then processed by a modulation filter-bank consisted of 9 modulation filters, from f_m = 1 Hz to f_m = 256 Hz, logarithmically spaced with a constant quality factor of Q =1(14,12,20). The absolute threshold for modulation detection was incorporated in the power spectrum calculation modeled by a -27 dB internal noise at the output of the filters (12).

2.2 Back-end: Decision device

Once the "internal representations" of the three intervals have been obtained, two additional outputs are needed: 1) the power in each auditory band, Ps_g and 2) the envelope power of each individual modulation filter, $Penv_{g,m}$. Finally the I2IR is calculated providing a map of the cues that the subject may use to identify the target among the three intervals. The decision device includes a "sensitivity" parameter that controls the minimum difference that the model is able to perceive. According to the Weber's law, this difference limen was assumed to be 1 dB I2IR.

The decision device will choose the interval that offers the highest I2IR that is defined by expression 1:

$$\mathrm{I2IRx}_{g,m}^{(i,j)} = 10\log\frac{\mathrm{Px}_{g,m}^{(i)}}{\mathrm{Px}_{g,m}^{(j)}} \tag{1}$$

The I2IR quantifies the power ratio between the intervals 'i' and 'j' both in spectral (PSM) and envelope (EPSM) domains. However, the integration of the cues across auditory and modulation channels differs. While in the envelope domain, all the I2IRs are taken into account by averaging all the quantities (expression 2), in the PSM, only the difference between maximum and minimum values is used in the decision device (expression 3). The $I2IR_s$ was tested following the procedure suggested in (11) and a free parameter φ was empirically fitted to the results at 1c/o in order to have 1 dB I2IRs at the estimated thresholds.

$$\mathrm{I2IR}_{env}^{(i,j)} = \frac{1}{GM}\sum_{1}^{M}\sum_{1}^{G}\left|\mathrm{I2IR}_{envg,m}^{(i,j)}\right|, \tag{2}$$

$$\mathrm{I2IR}_{S}^{(i,j)} = \varphi\max\left\{\ \mathrm{I2IR}_{Sg}^{(i,j)}\right\} - \min\left\{\ \mathrm{I2IR}_{Sg}^{(i,j)}\right\}. \tag{3}$$

Finally, the total I2IR is calculated using the sum of the envelope and spectral power differences.

$$\mathrm{I2IR}^{(i,j)} = \left(\alpha\ \mathrm{I2IR}_{env}^{(i,j)} + (1-\alpha)\mathrm{I2IR}_{S}^{(i,j)}\right) \tag{5}$$

The parameter α controls for the proportion of envelope / spectral I2IRs that the model uses to quantify the dissimilarity between the two intervals. α values ranges from 0 to 1.

Overall, the interval chosen by the decision device will be the one that exhibits the most salient differences. Nonetheless, a sensitivity factor (Θ) was included here reflecting the perceptual limits of the auditory system. In accordance to the Weber's law, this sensitivity factor was set at 1 dB interval-to-interval ratio for $\alpha = 1$, which was able to reproduce the experimental results from (12). For conditions where $\alpha < 1$, the sensitivity factor varied accordingly ($\Theta = \alpha$).

3. Methods

3.1 Stimuli generation and equipment

All psychoacoustical tasks were carried out using the AFC framework implemented in MATLAB (22). The stimuli were generated at a sampling frequency of 44100 Hz and converted to analogue signals using an RME Fireface sound card. The resulting signal was amplified (SPL headphones amplifier) and presented to the listener through Sennheiser HD650 headphones. The experiments were performed in a double-walled sound-attenuating booth.

The ripple stimuli were produced similarly as in (1,23). The mathematical description of the stimulus is characterized by:

$$S_i(x_i t) = A\sin(2\pi f_{c_i} t + \gamma_i)(1 + m\sin[2\pi(f_m t + \Omega x_i + \phi)]), \quad (6)$$

For the temporal modulation, the sinusoidal carrier is modulated in amplitude, where m is the modulation depth and f_m the modulation frequency. In the case of spectro-temporal modulation, Ω is the spectral ripple density and x_i the instantaneous space-frequency related to the center frequency of the octave bands $x_i = \log_2(f_{ci}/f_{cb})$. For spectral modulation m = 0, it follows:

$$S_i(x_i t) = 10^{C/2(\sin(2\pi(x_i f_c + \theta_0))/20} A\sin(2\pi f_{c_i} t + \gamma_i), \quad (7)$$

where C is the spectral contrast that controls the modulation depth in the spectral domain. The stimuli were generated in the frequency domain as the sum of 256 equal-amplitude carrier tones per band, logarithmically spaced. The phase of all the carriers was randomized. Sinusoidal AM was applied by additional sidebands placed at $f_{ci} \pm f_m$ with instantaneous phases increasing according to the frequency space x_i. The two conditions included in the present study were found to be the most significant combinations of spectral density and modulation frequencies for the narrowband STM sensitivity experiment of Mehraei et al. (2014). These are 1000 Hz, f_m = 4 Hz, Ω = 2 c/o and 4000 Hz, f_m = 4 Hz, Ω = 4 c/o.

3.2 Procedure and listeners

All psychoacoustical experiments were measured at 35 dB sensation level (SL). Two unmodulated 1-octave band noises, centered at 1 and 4 kHz were used to estimate auditory thresholds. Then, in each of the tasks, the listeners had to identify which interval contained the deviant stimulus in a 3-interval AFC paradigm. In the initial condition, the target signal was clearly identifiable whereas the other two intervals contained unmodulated noise. The adaptive tracking procedure of 1-up 2-down approximated the 70% point on the psychometric function (24). Listeners were presented with three runs per condition. If the measured thresholds differed more than 3 dB, a fourth threshold was performed. Fifteen subjects participated in the experiment; they were all students of different nationalities, ranged between 23 and 26 years with a median of 24.5 years. Their audiometric thresholds were below 20 dB hearing level (HL) for the explored frequencies.

4. Experiment I: Modulation detection

4.1 Method

For measuring the TM and the STM detection thresholds, the modulation depth was varied in dB (20log(m)). The starting modulation depth of 0 dB was decreased in steps of 6 dB. After the first reversal, the step size decreased by 4 dB. Finally, the mean of 6 reversals using steps of 2 dB were used to estimate the threshold. Likewise, the SM detection thresholds were estimated by varying their spectral contrast C in dB. The considered fully modulated condition was 30 dB peak-to-valley in the spectral domain. The results were then presented in terms of the difference between the SMD threshold and the initial condition for a fair comparison with the other types of modulations.

4.2 Results & Discussion

Figure 3 shows the data obtained in the proposed detection tasks. The individual data, as well as the boxplots are presented together with the model simulations for identical tasks. For each center frequency, the thresholds for the spectro-temporal (STMD) and only spectral conditions (SMD) tend to be lower than are for the temporal condition (TMD). This suggests that STMD represented an easier task compared to TMD as was also observed in (7,10).

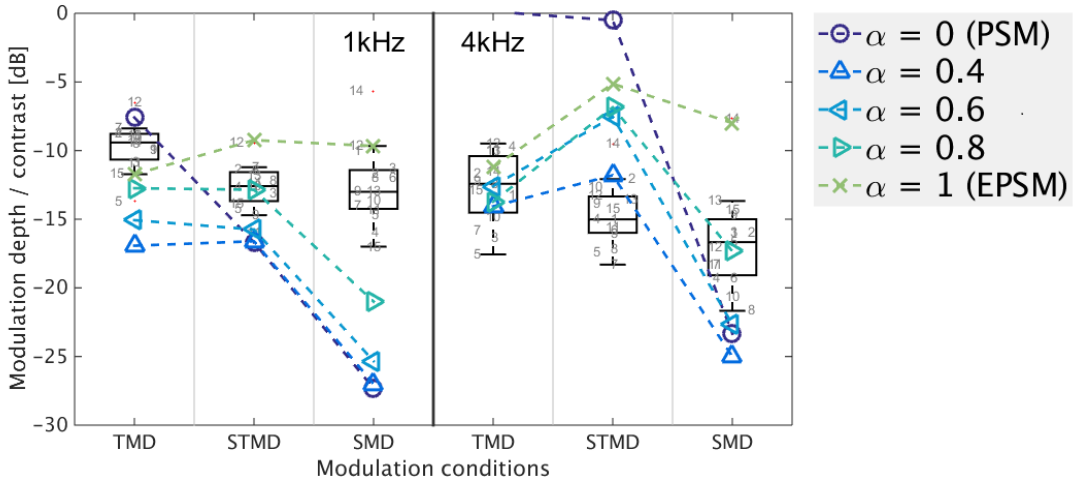

Figure 3 - Detection tasks for temporal, spectral and spectro-temporal modulations. Results correspond to NH listeners and the basic auditory model using different α values. The results showed similar trends but thresholds were overestimated, especially at 1 kHz.

In a recent study (25), TMD, SMD and STMD were measured in normal hearing, hearing impaired listeners and cochlear implantees. Overall, their results for NH differed from the ones presented here in the sense that the STM sensitivity presented more elevated thresholds than the TM. However, the method used for both threshold measurements and the stimuli generation were different. While here, both stimuli were generated in the same way and the only difference was the phase relationship of the sidebands, Won et al. (25) used a wideband noise carrier in the TMD. Therefore, it is more likely that the decreasing thresholds observed in the present study correspond to the use of additional cues besides spectral and temporal alone as stated in (7).

5. Experiment II: Modulation discrimination

5.1 Method

Modulation discrimination tasks were divided in two groups: 1) ripple discrimination and 2) modulation discrimination threshold. The spectral ripple discrimination (SRD) experiment provides an estimation of the maximum spectral density where the listener can distinguish between a spectral ripple with C = 30 dB and other ripples where the peaks and valleys are reversed, as in (2,3). For the spectro-temporal ripple discrimination (STRD), the listeners had to distinguish between an upward

and a downward fully modulated ripple. In contrast, in the case of the modulation discrimination threshold estimation, the target was a ripple fully modulated at low spectral densities. Whereas in the spectral discrimination threshold (SDT) experiment, the non-target intervals were unmodulated noise, the stimuli were temporally modulated with the same modulation frequency in the spectro-temporal discrimination threshold (STDT). In both cases, the task was to identify the spectrally modulated interval by decreasing the spectral density. For all the discrimination tasks, the starting spectral density was 1 c/o and this was varied in dBs (20log(Ω)) by increasing (ripple discrimination) or decreasing (modulation discrimination) the density in steps of 6 dB until the first reversal, 2 until the second reversal and 1 dB along the last 6 reversals.

5.2 Results & Discussion

Figure 4 shows the data for the two groups of discrimination tasks together, the left panels depict the discrimination thresholds while the ripple discrimination experiments are presented in the right panels. It seems consistent that the mean of the STDTs and SDT at 4 kHz is in the range of 0.13-0.15 c/o. If it is assumed that auditory filters bandwidth is about one third octave, it would correspond to a half of the bandwidth of an auditory filter. However, the STDT at 4 kHz showed consistent results for the majority of the subjects at spectral densities around 0.1 and even below. When a spectral density is introduced, the energy in the envelope domain decreases such that the subjects were more sensitive to this variation.

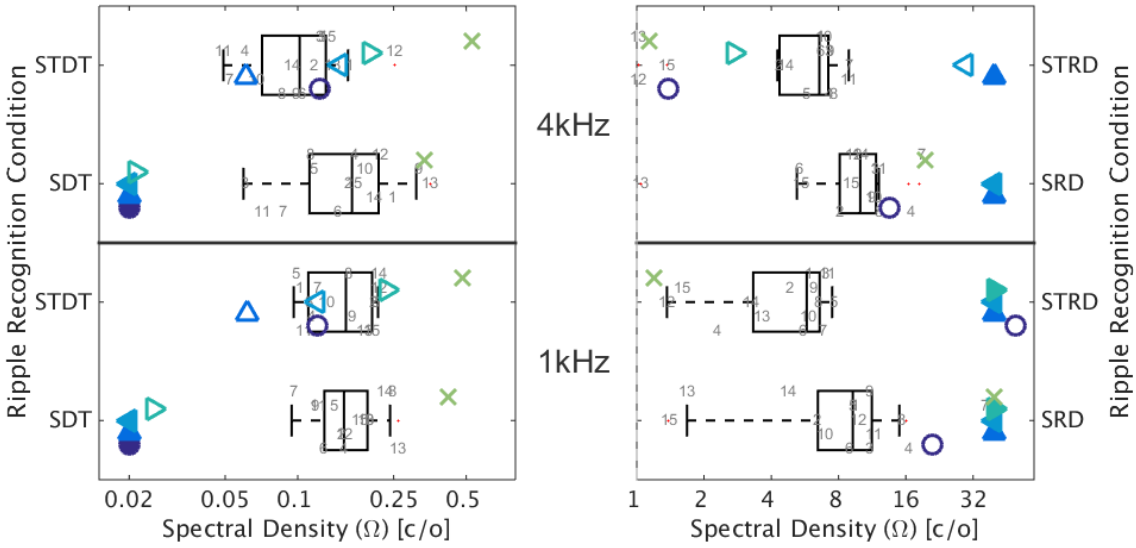

Figure 4 - Discrimination tasks, human data with model simulations. On the left, spectro-temporal discrimination threshold (STDT) as the minimun spectral density required to distinguish between TM and STM. Spectral ripple discrimination threshold (SDT) as the minimun Ω needed to identify a SM. Spectro-temporal ripple discrimination (STDT), maximun Ω for discriminating between upwards and backwards ST ripple. Spectral Ripple Discrimination (SRD) as in (3). Filled symbols showed the conditions where the procedure was skipped and the thresholds were overestimated.

Unlike the results of previous studies (15), where the mean SRD was 4.84 c/o, the data showed in Figure 4 showed that SRD relied in the range between 6 and 12 c/o with mean of 10 c/o. One can ascribe this better performance to the fact that the stimuli of the present study were bandlimited (1-octave) compared to the ones from (2). However, other essential difference is the presentation level. Whereas Henry et al. (2) presented the broadband spectral ripples at 65 dB SPL, the presentation level here was 35 dB SL, which for NH is much lower than in the previous study. Recently, Davies-Vem et al. (3) found also SDR around 7-8 c/o in NH when presenting the ripples at 55 dB SPL, which supports the idea that the presentation level may play a greater role than the bandwidth in the discrimination of the stimuli.

The discrimination task using ST ripples consisted of the discrimination between an upward and a downward ripple. As shown in (10), the modulation detection thresholds are affected by the direction of the ST ripple. However, Mehraei, et al. (7) did not find significant differences in STMD when using 1-octave band stimuli with different directions. Therefore, this opposition was proposed as an alternative to the SRD, where the amount of modulation, rate and density are the same and only the phase (direction) changes. The STRD limit presented here was in the range between 1 and 8 c/o with a mean of 5.13 c/o. The variance observed and the number of outliers suggested that this task may require more training or a different procedure.

6. Experiment III: Temporal and spectral resolution

6.1 Method

Besides the modulation detection and discrimination tasks, temporal and spectral resolution tasks were considered as an outcome measure related to the spectro-temporal modulation perception. Gap detection thresholds (GDT) were estimated by using as a marker (stimulus that contains the silence gap) the unmodulated 1-octave band noise, as in section 3.2. A silence gap was placed in the middle of the marker. The starting gap was 30 ms, which was reduced in dB (10 log (gap/1ms)) by 6 dB for the first reversal and then reduced to a half for every reversal until 0.5 dB for the last 6 reversals.

Spectral resolution was estimated by measuring frequency discrimination thresholds (FDT). The central frequency of the 1 octave band noise was shifted to a higher frequency in the target interval. The initial difference was 25%. The procedure was tracked in dB (20log(%)) with a final step size of 0.5 dB.

6.2 Results & Discussion

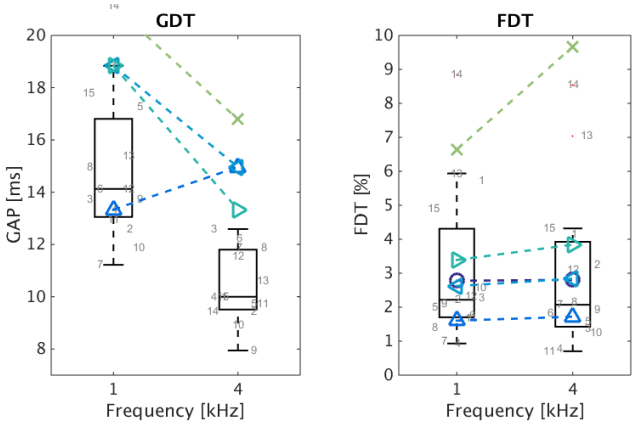

Figure 5 - Spectral and temporal resolution tasks. Gap detection thresholds (GDT) obtained by the model follow the trend of the NH results for the higher values of α but are overestimated. Frequency discrimination thresholds (FDT) were well reproduced by the model by using only spectral cues (α=0) and for the lower values of spectral-temporal combinations (α < 0.8).

The data from the temporal and spectral resolution tasks are showed in Figure 5. In this case the model simulations showed a clear change in trend between low and high α-values when simulating GDT. Whereas a greater contribution of the spectral cues showed lower GDTs at 1 kHz than at 4 kHz, a greater contribution of the temporal cues provided a trend, in line with the human data, but quite elevated. On the other hand, FDT mean results were fairly well reproduced by the model for all the α-values but for the EPSM alone.

7. Discussion

7.1 Analysis of the model simulations

The auditory-filter model was able to reproduce the TMD and SMD thresholds for different values of α. The best fit with the mean of the human data was found between α = 0.6 and α = 0.8. These two versions of the model were tested in order to reproduce T-MTFs and S-MTFs. The simulations could reproduce successfully the T-MTFs and shape of S-MTFs but shifted to lower spectral densities (Figure 6). However, the model was not able to capture STMD thresholds which were equal or higher than TMD especially at 4 kHz. This can be because the model only uses TFS information below 1.5 kHz and the temporal modulation may lead to some differences in the power spectrum. On the other hand, the EPSM alone (α = 1) overestimates the thresholds, not only for SMD, but also for STMD (Figure 3).

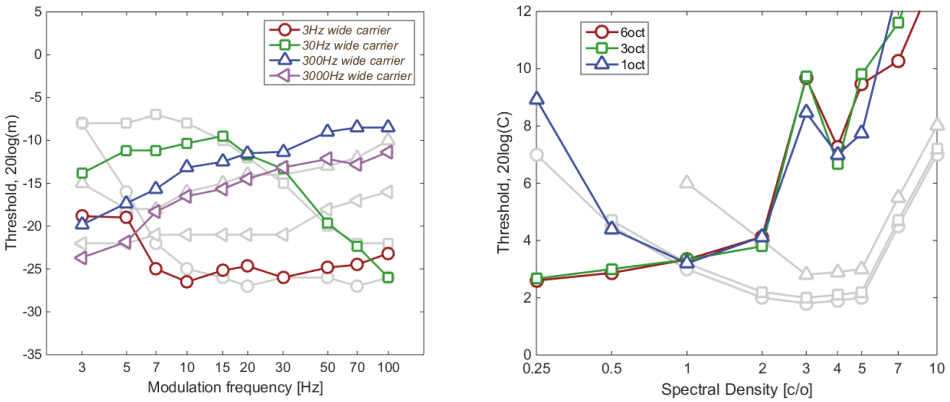

Figure 6 – T-MTFs and S-MTFs for broadband and narrowband carriers. Model simulations (α = 0.6 version). Simulations are compared to the data from (11,16).

As shown in Figure 4, model simulations were able to capture the STDT for combinations of PSM+EPSM (0.4 < α < 0.6) and PSM (α = 0) alone but not for EPSM (α = 1). This may suggest that the cues used in the discrimination of these stimuli are actually spectral rather than envelope based. Nevertheless, model failed in reproducing the SDT at both frequencies and thresholds were located well below 0.1 c/o. This suggests that some limitations for perceiving the spectral changes have not been taken into account. It would be of interest to understand, why the model fitted quite accurately to the human data when the noise is amplitude modulated (STDT) but not for SDT, therefore, further simulations including an internal noise in the auditory filters may provide more suitable simulations in both tasks.

The simulations of the ripple discrimination tasks showed that the model overestimated the SRD and STRD in the most of the conditions. As stated before, the purpose of the STRD test was to include a task where long-term power spectra and envelope power spectra should be similar so only combined spectro-temporal pattern differs. Therefore, a power-based model would not be expected to discriminate between them. Nevertheless, the model over-performed and, only in the cases of either PSM (α = 0) or EPSM (α = 1) alone, the model underestimated the results.

The different model versions were fitted by only one parameter (φ) in the condition (α = 0) and the sensitivity was adjusted to (Θ = α) in order to fulfill the Weber's law for T-MTFs. However, the adjustment of these two parameters is not completely independent and may be connected by a task that involves discrimination in both domains such as SRD or STRD.

7.2 Auditory-filter-model based vs STRF

The present model was able to reproduce temporal and spectral modulation detection, discrimination between temporal and spectro-temporal modulations as well as measures of temporal and spectral resolution. These simulations were obtained by means the combination of PSM and EPSM approaches and only one parameter was empirically fitted to the data. However, the model overestimated the ripple discrimination and underestimated the STM detection thresholds. One can

then discuss whether there were some features of the stimuli that were not captured by using a power-based metric. In the case of STMD and STRD, the task may involve the perception, not only of the differences in power, but also other features that may be crucial in the pattern recognition of complex STM. Figure 6, shows the visual representation of the stimuli used in the present study and explains how the STM do not provide a characteristic representation either in spectral or envelope power domains.

Overall, the simulations where the model fails could be due to 1) the power-based metric that may be substituted by an correlation-based optimal detector (26) or a temporal coherence, 2) the need of an across-channel processing stage as suggested in (6,27), 3) the need for further stages as adaptation or non-linear auditory filters (18,26), or 4) the need for other analysis for extracting an internal representation as suggested in models based on spectro-temporal receptive fields (4,10,27).

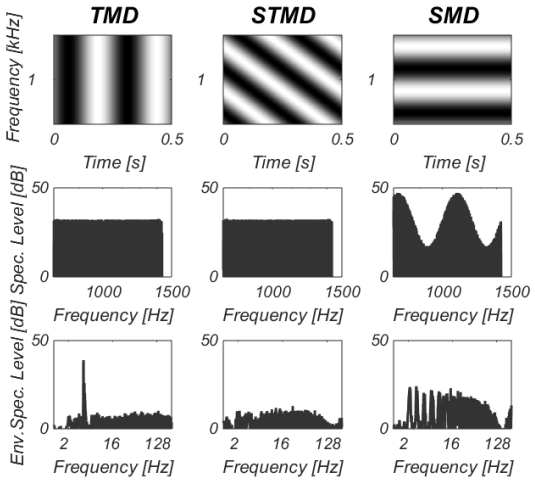

Figure 7 – Visual representation of the stimuli used in the detection tasks. First row shows the spectrograms of the TM, STM and SM stimuli. The spectrum of the three stimuli can be visualized in the second row, where the SM can be clearly identified. The envelope power spectrum is illustrated in the bottom row. While TM and SM present harmonic components in the spectrum, STM does not provide a characteristic representation in any of both domains.

STRF were used in (4) as a final cortical analysis preceded by an auditory model. When reproducing HI data, the model was fitted to the individual data making use of data psychoacoustical experiments such as auditory filter bandwidth, peripheral compression and STM sensitivity. As a result, a non-linear model fitted to individual data did not provide sufficient benefit and a simpler linear version that only used audiometric thresholds and the STMD was able to represent the variability of the STM data. This suggested that a model that analyzed the stimuli in terms of STRFs may not need a detailed front-end. However, the auditory-filter-model approach pursues the examination of effective auditory processing at different stages and their perceptual consequences.

An efficient model should be able to reproduce the perceptual consequences of the impairment of different stages of the auditory system. In that sense, the present approach should include stages that account for the reduction of frequency and temporal resolution as well as a back-end able to account for the discrimination of different spectro-temporal patterns.

8. CONCLUSIONS

The main findings observed in the present study are:

- The model based theories of auditory processing and perception, which only is fitted by one parameter, was able to reproduce several tasks related to spectral and temporal perception. The model simulations showed that the best combination of spectral and temporal cues was for $0.4 < \alpha < 0.6$.
- Experimental results showed better sensitivity for spectral and spectro-temporal modulation than for temporal modulations. However, all the different versions of the model underestimated the discrimination of spectro-temporal ripples, most likely because additional cues, besides purely spectral or temporal, have to be taken into account.
- The model overestimated in most of the discrimination tasks. Further stages in order to reproduce the perceptual limitations should be considered in the model.
- An efficient model that reproduces human perception by means of auditory processing should involve stages that can reflect specific impairments. A different back-end based on correlation or coherence may provide more suitable results in the discrimination tasks rather than spectro-temporal receptive fields.

ACKNOWLEDGEMENTS

This research was supported by the Centre for Applied Hearing research (CAHR). We thank M. Feręczkowski, J. Zaar, M.L. Jepsen, G. Mehraei and T. Biberger for helpful discussions.

REFERENCES

1. Bernstein JGW, Mehraei G, Shamma S, Gallun FJ, Theodoroff SM, Leek MR. Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. J Am Acad Audiol . 2013;24(4):293–306.
2. Henry B a, Turner CW, Behrens A. Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. J Acoust Soc Am . 2005;118(2):1111–21.
3. Davies-Venn E, Nelson P, Souza P. Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearinga). J Acoust Soc Am . 2015;138(1):492–503.
4. Bernstein JGW, Summers V, Grassi E, Grant KW. Auditory models of suprathreshold distortion and speech intelligibility in persons with impaired hearing. J Am Acad Audiol . 2013;24(4):307–28.
5. Elhilali M, Chi T, Shamma S a. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Commun. 2003;41(2-3):331–48.
6. Chabot-Leclerc A, Jørgensen S, Dau T. The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction. J Acoust Soc Am . 2014;135(6):3502–12.
7. Mehraei G, Gallun FJ, Leek MR, Bernstein JGW. Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. J Acoust Soc Am . 2014;136(1):301–16.
8. Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am. 1979;66:1364.
9. Green DM. “Frequency” and the Detection of Spectral Shape Change. In: Moore BCJ, Patterson RD, editors. Auditory Frequency Selectivity . Boston, MA: Springer US; 1986. p. 351–9.
10. Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. J Acoust Soc Am . 1999;106(5):2719–32.
11. Eddins D a, Bero EM. Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. J Acoust Soc Am. 2007;121(1):363–72.
12. Dau T, Kollmeier B, Kohlrausch A. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am. 1997;102:2906.
13. Dau T, Verhey J, Kohlrausch A. Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. J Acoust Soc Am. 1999;106(5):2752–60.
14. Jørgensen S, Dau T. Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. J Acoust Soc Am. 2011;130(September 2011):1475.
15. Patterson, R. D. & Moore BCJ. Auditory filters and excitation patterns as representations of frequency resolution. Frequency selectivity in hearing. 1986. p. 123–77.
16. Dau T, Kollmeier B, Kohlrausch A. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J Acoust Soc Am. 1997;102:2892.

17. Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. JARO - J Assoc Res Otolaryngol. 2007;8(3):384–92.
18. Dau T, Püschel D, Kohlrausch A. A quantitative model of the "'effective'"signal processing in the auditory system. I. Model structure. J Acoust Soc Am. 1996;99:3615.
19. Patterson RD, Nimmo-Smith I, Holdsworth J, Rice P. An efficient auditory filterbank based on the gammatone function. APU Rep. 1988;2341.
20. Ewert SD, Dau T. Characterizing frequency selectivity for envelope fluctuations. J Acoust Soc Am. 2000;108(3 Pt 1):1181–96.
21. Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. Hear Res. Amsterdam,: Elsevier; 1990;47(1-2):103–38.
22. Ewert S. AFC - A modular framework for running psychoacoustic experiments and computational perception models. Proceedings of the International Conference on Acoustics AIA-DAGA 2013 . Merano, Italy; 2013. p. 1326–9.
23. Litvak LM, Spahr AJ, Saoji A a, Fridman GY. Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. J Acoust Soc Am. 2007;122(2):982–91.
24. Levitt H. Transformed Up-Down Methods in Psychoacoustics. J Acoust Soc Am. 1971;467–77.
25. Won JH, Moon IJ, Jin S, Park H, Woo J, Cho YS, et al. Spectrotemporal modulation detection and speech perception by cochlear implant users. PLoS One . 2015;10(10):1–24.
26. Jepsen ML, Ewert SD, Dau T. A computational model of human auditory signal processing and perception. J Acoust Soc Am . ASA; 2008;124(1):422–38.
27. Schädler MR, Warzybok A, Ewert SD, Kollmeier B. A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. J Acoust Soc Am. 2016;139(5):2708–22.