



COVER PAGE

Document downloaded by @DAEL

Mon Oct 26 14:02:19 2026

For personal use

When automatic English translation is provided, only the original document is authentic.

The EAA cannot be held responsible of any translation error

Bibliographical reference

Attempt to Predict Temporal Modulation Transfer Function by Amplitude Modulation Following Responses, Jing Chen, Zhen Fu, Jiping Wu and Xihong Wu, *Acta Acustica* **vol. 104** (Number 5), 2018, pp. 821-824

DOI

<https://doi.org/10.3813/AAA.919237>

Attempt to Predict Temporal Modulation Transfer Function by Amplitude Modulation Following Responses

Jing Chen, Zhen Fu, Jiping Wu, Xihong Wu

Department of Machine Intelligence, Speech and Hearing Research Center, and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100081, China. chenj@cis.pku.edu.cn

Summary

The temporal modulation transfer function (TMTF) reflects the temporal resolution of the auditory system. The measurement of TMTF is based on a psychoacoustic paradigm that is not applicable in the clinic. Previous studies addressed that amplitude modulation following responses (AMFRs) could be used to objectively evaluate listeners' temporal auditory acuity, but no existing method could predict TMTF from AMFRs. The present study proposes a method as an attempt to do so. A logistic function was derived from the simulation of AMFRs on human scalp. To evaluate the new method, both AMFRs and behavioral TMTF experiments were conducted with 12 normal hearing listeners. The test showed that the predicted thresholds were correlated with, but deviated systematically from the behavioral thresholds at four modulation rates (32, 64, 128 and 256 Hz). Limitations of the method are discussed.

© 2018 The Author(s). Published by S. Hirzel Verlag · EAA. This is an open access article under the terms of the Creative Commons Attribution (CC BY 4.0) license (<https://creativecommons.org/licenses/by/4.0/>).

PACS no. 43.60.-c, 43.64.-q, 43.66.-x, 43.71.-k

1. Introduction

Sensitivity to fluctuations of temporal envelope of sounds is not only a fundamental property of the auditory system, but it is also an important predictor of speech perception for hearing impaired-listeners [1]. The two dominant ways to measure a listener's sensitivity to envelope fluctuations are to determine how sensitive they are to sinusoidal modulations in the amplitude of a sound (the temporal modulation transfer function, TMTF), or how sensitive they are to a gap in a continuous sound or between two sounds (gap detection). It has been reported that the TMTF could be informative about the ability of a listener to detect a gap [2]. Since the measurement of TMTF is based on a behavioral paradigm, it is difficult to test subjects who are unable to give accurate behavioral responses (e.g., young children, individuals with language impairment or cognitive difficulties). Therefore, developing an objective and efficient measurement to assess TMTF is of clinical interest. The present study addresses this need.

In previous studies, the objective amplitude modulation following responses (AMFRs) were used to evaluate temporal auditory acuity. Purcell *et al.* [2] recorded AMFRs evoked by white noise that was amplitude modulated at

25% modulation depth with a sweep of modulation frequencies from 20 to 600 Hz. They found that the frequency at which the AMFR was no longer detectable was significantly correlated with behavioral measurements of the maximum perceptible modulation frequency, gap detection and pure tone thresholds. Luke *et al.* [3] measured electrically evoked AMFRs and behavioral modulation detection thresholds (MDT) for cochlear-implant listeners, and found that AMFR amplitudes and signal-to-noise ratios for 40 Hz were significantly related to the MDTs for 20 Hz. These results indicate that AMFRs may be used as an objective measure of temporal sensitivity. However, no existing methods can be used to predict MDT from AMFRs. The present study proposes a method as an attempt to predict the TMTF via analysis of AMFR strength as a function of the modulation depth, and tests the effectiveness of this method in normal-hearing listeners.

2. Prediction method

Two main factors are usually manipulated for AMFR stimuli: modulation frequency (f_m) and modulation depth (MD). The strength of AMFRs (measured in relative amplitude, RA) can be represented as a function of MD for each f_m . Possibly, the MDT can be predicted for each f_m according to the RA-MD function of AMFRs. However, since the sampling of MD in recorded data is relatively limited, it is difficult to decide on the mathematical form

Received 2 March 2018,
accepted 31 August 2018.

of the RA-MD function. Hence, a computational model was employed to generate simulated AMFRs, with which the functional form was determined.

2.1. Simulation of AMFRs

Figure 1 shows the block diagram of the model used in the present study for simulating AMFRs. The model combines a model of auditory peripheral and central responses [4] with the convolution model of frequency-following responses [5]. The stimuli were first processed through auditory-nerve (AN) models with different center frequencies (CFs). The discharge rates of AN fibers were fed into higher processing centers: ventral cochlear nucleus and inferior colliculus (IC), connecting by a synapse model. The discharge rates of IC fibers were summed across all channels and convolved with a given unitary response (UR) function to generate the simulated AMFRs. A set of 500 CFs ranging from 0.1 to 10 kHz was used.

2.2. Calculation of threshold

The simulated AMFRs were used to facilitate determination of the form of the fitting function. The input signal of the model was sinusoidal amplitude modulated (SAM) white noise, as given by Equation (1)

$$A(t) = (1 + m \sin 2\pi f_m t) N, \quad (1)$$

where $A(t)$ represents the signal waveform as a function of time t , m corresponds to the MD and N represents the carrier of white noise. The signal bandwidth is 10 kHz and sound level is 65 dB SPL. There were in total 31 MDs from -30 dB to 0 dB in 1-dB steps. Due to the stochastic characteristic of white noise as the carrier, AMFR was simulated 100 times for each MD condition, and the carrier was randomly produced for each simulation. Then the RA was calculated to represent the strength of AMFRs. The AMFR was first transformed into its magnitude spectrum. Then, the averaged amplitude of the bins near the target frequency relative to the amplitude of a wider surrounding region was used as a measure of RA. An RA-MD function based on simulated AMFRs is shown in Figure 2. The function grows slowly at the lower and greater depth sides while rapidly at moderate depth, and the function almost reached saturation at the greatest depth. In consideration of this typical S-shaped characteristic, the logistic function was chosen as the fitting function. The logistic function used in the present study is given by Equation (2),

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}, \quad (2)$$

where x represents MD, L represents the maximum value of RA, k represents the slope and x_0 represents the MD corresponding to $L/2$. This value of x_0 also corresponds to the MD where the slope is the greatest. Hence we assumed x_0 to be the predicted MDT.

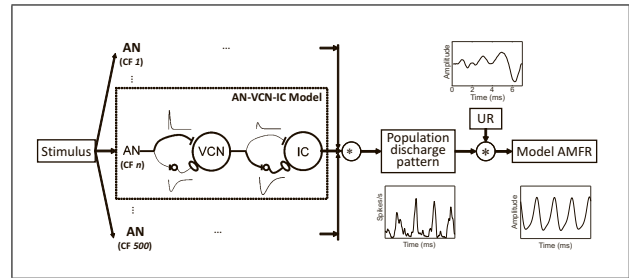


Figure 1. Architecture of the model used to simulate AMFRs. Temporal patterns of population discharge, UR and simulated AMFR are also illustrated.

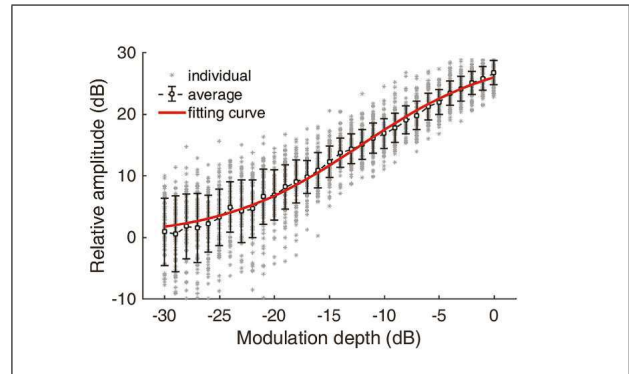


Figure 2. (Colour online) RA of simulated AMFRs as a function of MD for 32 Hz f_m (red line). Individual (gray scatter) and averaged (black circles) RA are also shown. The error bar indicates one standard deviation.

3. Experiments

3.1. Participants

Twelve normal-hearing listeners (6 males, mean age = 22.2 years) participated in both TMTF and AMFRs experiments. All were Mandarin-native speakers. All gave informed consent and were paid.

3.2. Behavioral measurement of TMTF

Behavioral TMTF was measured for all subjects in a sound-attenuating booth. A custom MATLAB program was used to present stimuli and collect responses. The stimuli were SAM noise with 48-kHz sampling rate and 1.5-s duration, including 25-ms rise-fall ramps. Signals were presented binaurally at 65 dB SPL through a Sennheiser HD 265 headphone, and signals' levels were equalized based on RMS.

There were 8 f_m conditions in this experiment: 4, 8, 16, 32, 64, 128, 256 and 512 Hz. A three-alternative forced choice (3AFC) procedure was used. For each trial, one SAM white noise and two unmodulated white noises were randomly assigned to the three intervals. The inter-stimulus interval was 0.5 s. The carriers were random for the three intervals in each trial. Subjects were instructed to identify the modulated interval and were provided with feedback. The MD was manipulated through a 3-down 1-up algorithm with an initial value of -5 dB. A reversal

was defined as occurring when the variation direction of depth changed. The step size was initially 8 dB, but was reduced by half after every two reversals until it reached 1 dB. For each f_m condition, the threshold was calculated as the mean of the last 4 reversals after 10 reversals were obtained. Each subject received a short practice session before the formal test.

3.3. AMFR experiment

3.3.1. Stimulation and recording

Stimuli were SAM noise, generated offline in MATLAB with 48 kHz sampling rate and stored for playback. Listeners sat in a double-walled sound-attenuating booth. Stimuli were presented binaurally at 65 dB SPL through an ER-3A insert earphone. EEG recordings and data storage were controlled by the NeuroScan Acquire 4.3 software with a SynAmps2 system. A vertical electrode montage was adopted, with three Ag-AgCl electrodes placed at Cz as non-inverting, right earlobe as inverting, and forehead as ground. The EEG recordings were amplified (gain of 20000), bandpass filtered (10–2000 Hz, 6 dB/octave) and digitized at 20-kHz sampling rate.

Values of f_m were 32, 64, 128 and 256 Hz, and MDs were -5, -10, -15, -20, -25 and -30 dB, which were determined from a pilot experiment to ensure AMFRs varied in a reasonable range. Hence there were 24 (4×6) conditions for each subject. The whole test was divided into 10 blocks. For each block, 24 pieces of SAM noise corresponding to the 24 conditions were presented in a random order. The duration of each SAM noise was 10 s with an inter-stimulus interval uniformly distributed between 2.5 and 3 s.

3.3.2. Data analysis

The data analysis was performed offline. The first and last 0.5 s of each 10-s continuous recording were discarded to avoid the onset and offset effect of neural responses. The remaining 9 s was segmented into segments of 125 ms in length, resulting in 72×10 segments for each condition. A single segment would be considered contaminated and was rejected if it contained voltages exceeding $\pm 25 \mu\text{V}$. Band-pass filtering was conducted on each artifact-free sweep to enhance the component corresponding to f_m . The passbands were 20–40, 40–80, 80–160 and 160–320 Hz for $f_m = 32, 64, 128$ and 256 Hz, respectively. After filtering, segments of the same condition were ensemble averaged to reduce noise. Then the RA was calculated to represent the strength of AMFRs.

4. Results

4.1. Behavioral measurement of TMTF

The average TMTF of all listeners (not shown) is similar to that in a previous report [6], confirming that the behavioral experiment was effective. The function was relatively flat for f_m up to 16 Hz, indicating that for low rates performance was limited by amplitude resolution rather than temporal resolution.

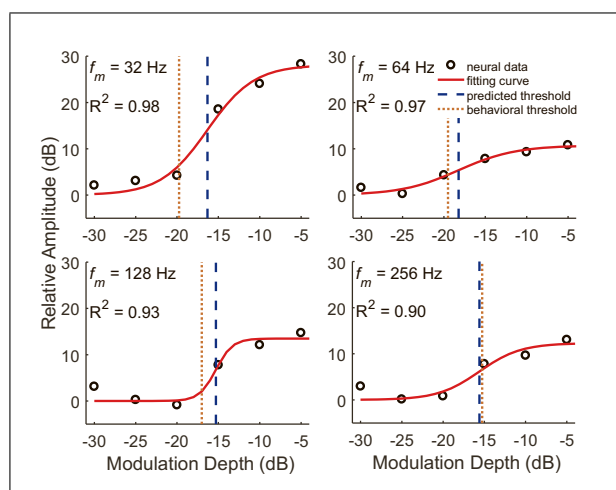


Figure 3. (Colour online) Results for the four f_m conditions for one subject. Black circles represent RA of each condition. Red curves represent the logistic fits. Measured and predicted MDT are given by the yellow dotted and blue dashed lines, respectively.

4.2. Objective prediction of MDT

Figure 3 shows the fitting result of one subject, in which the distribution of RA data (black circles) matched well with logistic functions (red curves) for all f_m . R^2 (goodness of fit) averaged across all fitting results of the 12 subjects was about 0.91 (± 0.07). The predicted MDTs (blue dashed lines) determined from the fitting function are shown for comparison with behavioral measured MDTs (yellow dotted lines). The differences between predicted and behavioral thresholds range from 0.2 to 3.5 dB. Figure 4 shows the predicted MDT as a function of the behavioral MDT of each listener for each f_m condition. Pearson's analysis indicated a significant correlation between predicted and behavioral thresholds for all f_m ($p \leq 0.028$). These results show that AMFRs might be useful for objectively predicting TMTF [2, 3]. There were no significant correlations between any two of the metrics: behavioral thresholds, the maximum RA, individual age and absolute thresholds ($p \geq 0.087$).

It should be noted that the threshold was likely to be overestimated. A two-way ANOVA (2 methods $\times$ 4 modulation frequencies) revealed that the main effects of method ($F_{1,11}=10.038$, $p<0.01$) and f_m ($F_{3,33}=51.221$, $p<0.001$) were both significant, and their interaction was also significant ($F_{3,33}=10.410$, $p<0.001$). Post hoc pairwise comparisons revealed that the predicted threshold was significantly higher than the behaviorally measured one ($p \leq 0.013$) for f_m of 32, 64 and 128 Hz, but the threshold difference was not significant for 256 Hz.

5. Discussion

TMTF measurement is an important method to evaluate the temporal resolution of the auditory system, but it depends on subjective responses. Previous studies found that

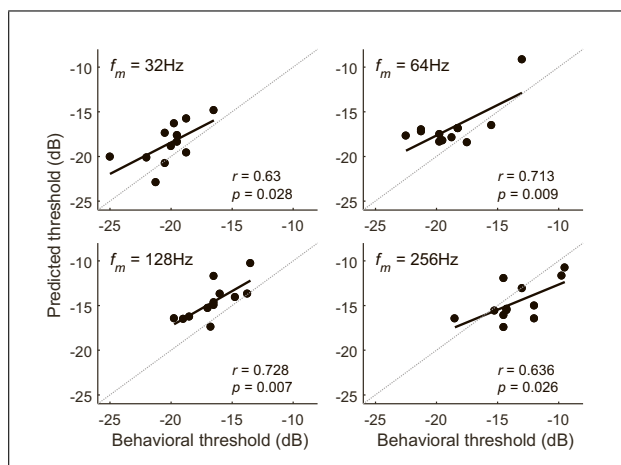


Figure 4. Correlation of predicted thresholds with behavioral thresholds across subjects, for four different f_m conditions.

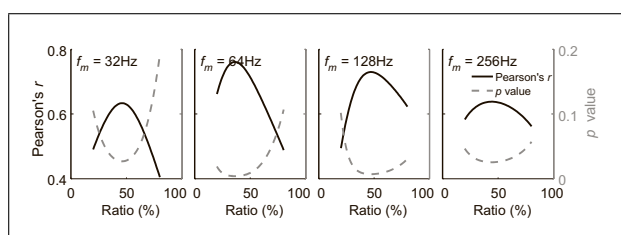


Figure 5. Pearson's r (solid lines) and p value (dashed lines) of correlation analyses as function of the ratio between the RA adopted for threshold prediction and the maximum RA, for four different f_m conditions.

AMFRs were highly correlated to some behavioral measures of temporal acuity [2, 3]. Here we investigated a method to predict the TMTF from logistic functions fitted to AMFRs recorded for different modulation rates. The results showed that for a given modulation rate the objectively predicted thresholds were correlated with behavioral thresholds across normal hearing listeners. However, the predictions deviated systematically from the observed thresholds and these deviations depended on the modulation rate. Thus, further work is required.

First, as the relative contribution of different sources in the auditory pathway to AMFRs likely depends on the modulation rate [7], the effect of different generators was not considered in the present study. The threshold was defined as x_0 of the logistic function in the current method, but a further analysis (see Figure 5) indicated that the accuracy of predictions might be improved by adopting different points on the function relating RA to MD (see Figure 3) as predicted thresholds for different f_m . To this end, Figure 5 shows, separately for each f_m , the correla-

tion (and the corresponding p value) between behavioral thresholds and a wide range of predicted thresholds as a function of the ratio between the RA value adopted for threshold prediction and the maximum RA. The results show that the best correlation occurs at different ratios for different f_m conditions, which might be due to the effect of different generators. Second, the validation tests were only conducted with normal-hearing listeners, the effectiveness of the method remains unclear for hearing-impaired listeners. A potential problem is that the fitting function was based on the auditory models of normal hearing. When it is used for impaired hearing, extra parameters or different functions should be introduced to account for the effect of hearing loss [4, 5]. Finally, other types of SAM signals should be tested in future work, since a carrier of narrow-band noise or pure tones could be more accurately controlled to investigate the temporal processing of the auditory system [8].

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61473008, 61771023, and 11590773).

References

- [1] G. A. Takahashi, S. P. Bacon, P. Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J. Speech Hear. Res.* **35** (1992) 1410–1421.
- [2] D. W. Purcell, S. M. John, B. A. Schneider, T. W. Picton: Human temporal auditory acuity as assessed by envelope following responses. *J. Acoust. Soc. Am.* **116** (2004) 3581–93.
- [3] R. Luke, D. L. Van, M. Hofmann, W. A. Van, J. Wouters: Assessing temporal modulation sensitivity using electrically evoked auditory steady state responses. *Hear. Res.* **324** (2015) 37–45.
- [4] P. C. Nelson, L. H. Carney: A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J. Acoust. Soc. Am.* **116** (2004) 2173–2186.
- [5] T. Dau: The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J. Acoust. Soc. Am.* **113** (2003) 936–950.
- [6] N. F. Viemeister: Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.* **66** (1979) 1364–1380.
- [7] P. Tichko, E. Skoe: Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hear. Res.* **348** (2017) 1–15.
- [8] T. Dau, B. Kollmeier, Armin Kohlrausch: Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.* **102** (1997) 2892–2905.