Monaural and binaural subjective modulation transfer functions in simple reverberation

ERIC R. THOMPSON AND TORSTEN DAU

Centre for Applied Hearing Research, Ørsted•DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark

The temporal intensity envelope of a signal is filtered by the transmission channel through which it passes. The amount of reduction for a given envelope, or modulation, frequency has been called the modulation transfer function (MTF) and can be derived from the impulse response of the transmission channel [Schroeder, M.R. (1981) Modulation transfer-functions: Definition and measurement, Acustica, 49, 179-182]. The envelope of a speech signal is critical for intelligibility, and the speech transmission index (STI) predicts the intelligibility of speech through a given transmission channel based on its MTF [Houtgast, T. and Steeneken, H.J.M. (1973) Modulation transfer-function in room acoustics as a predictor of speech intelligibility, Acustica, 28, 66-73]. In the present study, the results of intensity modulation detection experiments with broadband noise carriers are reported in monaural and binaural conditions, with single reflections at different arrival times in the two ears and with a simulated room impulse response. The monaural data describe a subjective MTF, which is similar to the physical MTF. Binaurally, the thresholds are consistently lower than the monaural thresholds, especially at frequencies where there is a large interaural modulation phase difference. These data show that binaural detection thresholds can be better than either ear alone and better than the predictions from either ear's physical MTF.

INTRODUCTION

The temporal amplitude envelope of a speech signal is critical for understanding the words that are spoken. In fact, words can be understood from the temporal envelope of speech imposed on noise carriers (Shannon *et al.*, 1995). When a signal is transmitted through a channel (e.g., a telephone line, or a room), the temporal dynamics of the signal are distorted by echoes and reverberation in the channel. In general, reverberation acts as a low-pass filter, attenuating fast modulations and allowing slow modulations to pass through unaffected. The amount of modulation transmitted through the channel as a function of the modulation frequency, or the modulation transfer function (MTF), can be derived from the impulse response (IR) of the channel. Given an IR $h(t)$, the MTF is the Fourier transform of the squared IR normalized by the total energy of the IR (Schroeder, 1981):

$$\mathrm{MTF}(f_m) = \frac{\int_0^\infty h^2(t)\, e^{-j2\pi f_m t}\, dt}{\int_0^\infty h^2(t)\, dt}, \qquad \text{(Eq. 1)}$$

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

where fm is the modulation frequency. This means that if a signal is transmitted through the system with an input intensity envelope given by

$$i_{in}(t) = i_0\left[1 + \sin(2\pi f_m t)\right], \qquad \text{(Eq. 2)}$$

with mean intensity i0, then the output intensity envelope will be

$$i_{out}(t) = i_r\left[1 + m\sin(2\pi f_m t + \phi)\right],\ m = \left|\mathrm{MTF}(f_m)\right|,\ \phi = \angle\mathrm{MTF}(f_m), \qquad \text{(Eq. 3)}$$

where ir is the output mean intensity, m is the resultant modulation depth (with values in the range [0,1]) and φ is the modulation phase shift.

Houtgast and Steeneken (1973) used the MTF as a predictor of the intelligibility of speech in rooms in their speech transmission index (STI). To calculate the STI, the MTF is measured (or calculated) in octave bands at several modulation frequencies that are deemed to be relevant for speech. The magnitudes of the MTF in dB are then weighted, added together, and scaled to produce a single value STI, which has been shown to correlate well with speech intelligibility scores in reverberation in many different tests (see, e.g., Houtgast & Steeneken, 1985). In the STI calculation from Houtgast and Steeneken (1985), only modulation frequencies from 0.63 to 12.5 Hz are included, citing these frequencies as the most critical for speech intelligibility. Modulation frequencies in this range are most affected by reflections that arrive later than 40 ms after the direct sound. This corresponds well with the room acoustics literature (e.g. Bradley *et al.*, 2003), where there is a temporal divide at 50 ms after the direct sound between early reflections, which are considered to enhance speech intelligibility, and late reflections, which have a negative effect on speech. Some recent evidence suggests, however, that modulation frequencies greater than 12 Hz are also important in the decoding of consonants (Christiansen and Greenberg, 2005). These sounds may be adversely affected by early reflections. Therefore, it may be of interest to investigate effects at modulation rates above the usual range of the STI calculation.

The STI has been a single-channel measurement, but humans listen with two ears, which can have large differences in MTF magnitude and phase from a single source. Those interaural differences in MTF arise due to the presence of the head between the ears, and can lead to perceptible interaural intensity fluctuations. If the interaural intensity difference (IID) is defined as the ratio of the ears' intensity envelopes as a function of time, then the IID resulting from the signal defined in eq. 2 is

$$\mathrm{IID}(f_m, t) = 10\log\left[\frac{i_{r,L}}{i_{r,R}}\frac{1 + m_L\sin(2\pi f_m t + \phi_L)}{1 + m_R\sin(2\pi f_m t + \phi_R)}\right], \qquad \text{(Eq. 4)}$$

where the subscripts L and R indicate the left and right ears, respectively. Miyata *et al.* (1991) measured speech intelligibility and a binaural subjective MTF in a reverberation chamber. From their data, they proposed that the overall binaural MTF was just the maximum of the two ears' MTFs at each modulation frequency, and that this "best ear" MTF could be used to calculate a binaural STI, which correlated well with their measured speech intelligibility data. However, while a reverberation chamber represents an extremely challenging listening environment with a long reverberation time and a diffuse sound field, it is not very representative of a typical room where speech is heard. Furthermore, it is expected that the diffuseness of the sound field in the reverberation chamber will eliminate any usable interaural differences that may give a cue for signal detection at levels below the best-ear thresholds.

The hypothesis in the present study was that simple reverberation, or even single reflections, could create large interaural modulation phase differences and perceivable interaural intensity fluctuations. These fluctuations could be used to create a real binaural advantage beyond just best-ear listening in the detection of intensity modulations.

METHODS

Psychoacoustic measurements were performed to determine the minimum modulation depth required to detect a sinusoidal intensity modulation imposed on a broadband noise carrier.

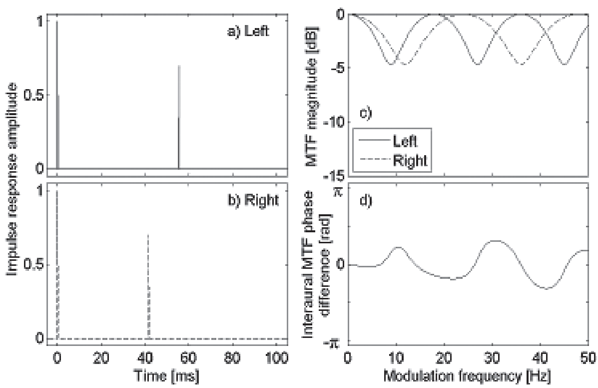

Fig. 1: Binaural impulse response amplitude with single reflections (a and b), MTF magnitude (c) and interaural modulation phase difference (d).

Procedure

A 3-interval, 3-alternative forced choice design was used with an adaptive 1-up, 2-down tracking rule for adjusting the modulation depth. With this tracking rule, the threshold represents the modulation depth at which the intensity modulation is correctly identified 70.7% of the time. Each run started with a modulation depth of -1 dB ($10\log_{10}m$) and an initial step-size of 2 dB. The step-size was halved after every second reversal until a final step-size of 0.5 dB was reached. The run continued for another six reversals, and the threshold was determined as the mean of these last six reversals. In order to avoid overmodulation of the stimuli, the modulation depth was limited to a maximum of 0 dB. If a listener could not successfully identify the signal interval on two successive trials with the maximum modulation depth, the track was skipped and no threshold was reported. Each listener repeated the measurements at each modulation frequency four times for each condition.

Stimuli

Broadband (0.1-5 kHz) pink-noise tokens were generated for each interval in each trial. In each trial, a sinusoidal intensity modulation was imposed on one randomly-selected interval according to eq. 5:

$$[1 + m\sin(2\pi f_m t + \phi)]^{0.5} x(t), \qquad \text{(Eq. 5)}$$

where m is the modulation depth, fm is the modulation frequency, φ is the start phase of the modulation and $x(t)$ is the amplitude of the noise carrier. The start phase, φ, was selected randomly for each trial from a uniform distribution over the range [0,2π]. The reference (unmodulated) and signal (modulated) stimuli were then convolved with a binaural impulse response, which was the direct sound only (anechoic), the direct sound plus a single reflection (see Fig. 1), or a simulated binaural room impulse response (BRIR) made with the ODEON room acoustics software (see Fig. 2). The room used for the simulation was a small classroom (6.7 m x 9.5 m x 3 m) with a reverberation time (T_{60}) of about 0.5 s. The relative level of the two ears' stimuli was maintained while the stimulus with the higher rms level was scaled to 65 dB SPL. Each stimulus was windowed to be 500 ms long, with 100 ms $\cos^2$ ramps. Measurements were then made with each ear individually (monoR and monoL) and binaurally with the anechoic and dichotic stimuli.

Figure 1 shows the IRs with single reflections at 55.6 ms after the direct sound in the left ear (panel a) and 41.7 ms in the right ear (panel b). With this reflection timing, the magnitude of the MTF has a minimum at 9 Hz in the left ear and at 12 Hz in the right ear (panel c), and an interaural modulation phase difference (IMPD) is created with a maximum of 0.73 rad at approx. 10.5 Hz. In order to sample the subjective MTFs across the two minima and at the maximum of the IMPD, measurements were made with this IR with signal modulations at 6, 9, 10.5, 12 and 15 Hz. In addition, a measurement was made with a 24 Hz modulation signal with the BRIR shown in Fig. 2. At this frequency, there is a larger IMPD (1.72 rad) than with the single reflection IR, and the MTF magnitude is relatively large (approx. -7 dB) in both ears.

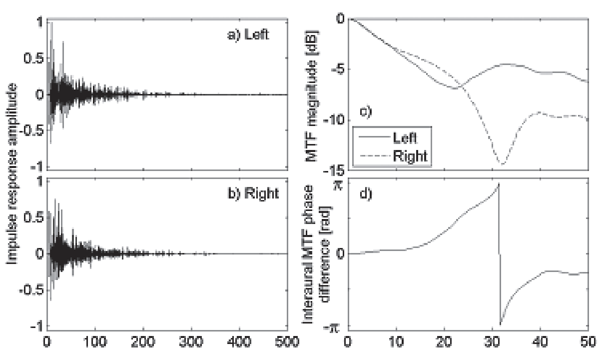

Fig. 2: Binaural room impulse response (BRIR) amplitude (a and b), MTF magnitude (c) and interaural modulation phase difference (d).

Equipment

The stimuli were generated on a PC using the AFC package for MATLAB (The MathWorks) and presented to the listener in a sound insulated booth through a sound card (RME DIGI96/8 PAD) and headphones (Sennheiser HD-580).

Listeners

Five test subjects participated in the experiment. They ranged in age from 24 to 32 years old, and had absolute thresholds for pure tones within normal audiometric limits (i.e., <15 dB HL). They were not paid directly for their participation, but all were directly affiliated with the research center, and included the first author of this paper.

RESULTS

The minimum modulation depths required to detect a sinusoidal intensity modulation imposed on a broadband noise carrier in the presence of single reflections are shown in Fig. 3a. The data points show the mean across test subjects and the standard error of the mean. In anechoic conditions (circles), the thresholds are fairly constant across the measured modulation frequencies with thresholds around -7 dB. This *intensity* modulation depth is comparable to a 20 dB *amplitude* modulation depth (see, e.g., Viemeister, 1979). The two monaural threshold curves with the single reflections show peaks at 9 and 12 Hz for the left and right ears, respectively. These curves correspond well with the MTFs from Fig. 1c, which show minima at the same modulation frequencies. The thresholds measured binaurally (squares) are consistently lower than or equal to either of the monaural thresholds, especially for the lower modulation frequencies.

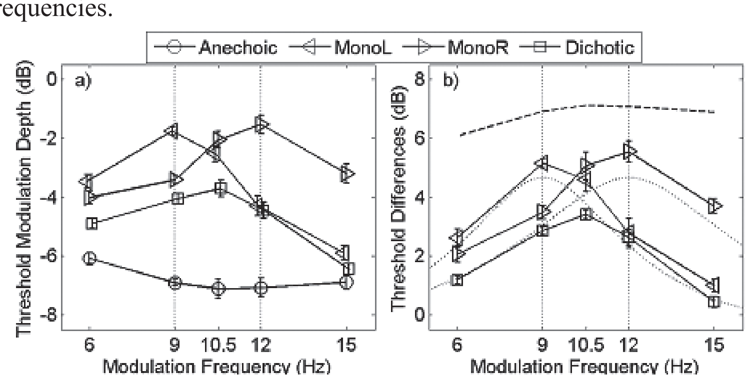

Fig. 3: Panel a) shows intensity modulation detection threshold depths with broadband noise carriers, for anechoic (circles), left and right ears alone with a single reflection (triangles pointing in the respective direction) and binaurally with single dichotic reflections (squares). The plots show the mean across test subjects and one standard error of the mean. Panel b) shows the differences in thresholds between the measurements with reflections and anechoic. The dotted lines show the inverse of the MTF.

When measuring with the BRIR shown in Fig. 2 and a 24 Hz modulation signal, the intensity modulations were undetectable for any of the test subjects when only listening with one ear (see Fig. 4), even with full modulation depth (m = 0 dB). However, by presenting the stimuli convolved with the BRIR, the listeners were able to use interaural differences to detect the intensity modulation. At the modulation frequency used with this BRIR (24 Hz), the anechoic threshold was around -7 dB, as with the lower frequencies measured earlier (see Fig. 3).

DISCUSSION

If a stimulus with a certain intensity modulation depth is transmitted through a room, the MTF can be used to calculate what the resulting modulation depth will be at the receiver. For example, if an input stimulus has m = -1 dB, and the magnitude of the MTF at that frequency is -5 dB, then the received stimulus will have a modulation depth of -6 dB. In theory, the minimum detectable modulation depth in reverberation for a single channel (i.e., monaural or diotic measurement) can be calculated as the difference between the anechoic threshold and the MTF in dB. The theoretical differences between the threshold with single reflections and the anechoic threshold are shown in Fig. 3b (dotted lines), along with the actual measured differences. The measured monaural difference curves (triangles) are close to their respective theoretical curves, but are always slightly larger than the theory predicts. In addition, the maximum possible threshold difference is shown in Fig. 3b with a dashed line. This maximum is defined here as the difference between the largest allowed modulation depth (0 dB) and the anechoic threshold at each modulation frequency.

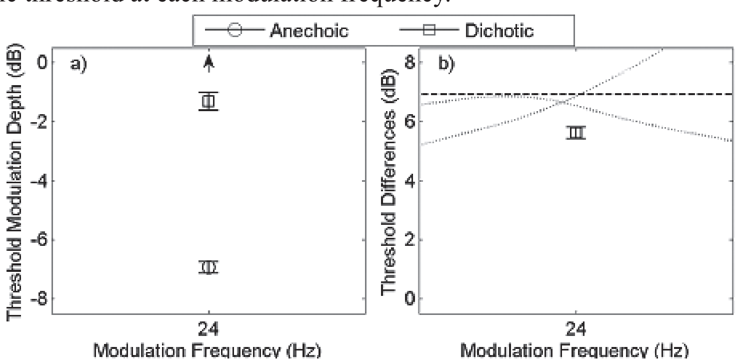

Fig. 4: Panel a) shows intensity modulation detection threshold depths with broadband noise carriers, for anechoic (circles) and binaurally with the BRIR (squares). The plots show the mean across test subjects and one standard error of the mean. The arrow at 0 dB indicates that the thresholds could not be measured monaurally. Panel b) shows the differences in thresholds between the measurements with reflections and anechoic. The dotted lines show the inverse of the MTF.

Miyata *et al.* (1991) suggested that the only advantage of binaural listening was to have two independent channels with the potential to extract the best information from either channel. According to that hypothesis, the thresholds measured binaurally with the dichotic single reflections should be the same as the lower of the two monaural thresholds. At most of the measured frequencies, this is true. However, at 10.5 Hz, where there is a local maximum in the interaural modulation phase difference, there is a significant difference between the binaural threshold (squares in Fig. 3a) and either of the monaural thresholds (triangles). This suggests that there can be a binaural advantage beyond just “best ear” listening, if there are usable interaural differences. The threshold differences in Fig. 3b show that while the monaural threshold differences were all above the theoretical difference curves, the binaural threshold differences are on or slightly below the theoretical curves. This shows that there can be a binaural advantage in a modulation detection task that is beyond bestear listening and beyond the prediction based on the physics of the stimulus.

The IR with single reflections is artificial in many respects, including that the reflections are perfect copies of the direct sound and that each reflection is only heard by one ear with no crossover to the other ear. Therefore, measurements were made with a more realistic IR from a simulation of the acoustics of a classroom. With this BRIR, the monaural MTFs were both at approximately -7 dB at the measured modulation frequency (24 Hz). With a threshold in anechoic conditions also at about -7 dB, the applied signal intensity modulation was undetectable for any of the listeners even with full modulation depth (0 dB). Once again, when listening with both ears, the intensity modulation was detectable at a modulation depth smaller than that predicted by the best-ear MTF. This suggests that this binaural advantage can also be used in real rooms.

The key component for this binaural advantage in intensity modulation detection is the interaural modulation phase difference (IMPD). This IMPD is created by interaural differences in the arrival time of reflections. The prevalence of these IMPDs in real rooms still needs to be investigated to determine if this binaural advantage can be expected in general, or if very specific acoustic conditions are required. In addition, it must be determined whether this binaural cue for intensity modulation detection can be used for speech processing.

CONCLUSIONS

Intensity modulation detection experiments were performed in anechoic conditions, with single reflections, monaurally and binaurally, and with a simulated room impulse response. These measurements showed that in anechoic conditions, the minimum intensity modulation depth for detection was about -7 dB for modulation frequencies from 6 to 24 Hz. The monaural subjective MTFs measured in the presence of single reflections were slightly larger than the predictions from the anechoic thresholds and the physical MTF derived from the IR. Listening with two ears to the stimuli with dichotic reflections provided a benefit that was beyond just bestear listening. This benefit was the result of interaural modulation phase differences that created dynamic interaural level differences. The applicability of this advantage to speech intelligibility needs to be investigated in further studies.

REFERENCES

Bradley, J. S., Sato, H., and Picard, M. (**2003**). "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244.

Christiansen, T. U., and Greenberg, S. (**2005**). "Frequency selective filtering of the modulation spectrum and its effect on consonant identification," 21st Danavox Symposium, Hearing Aid Fitting, edited by A. N. Rasmussen, and T. Poulsen.

Houtgast, T., and Steeneken, H. J. M. (**1973**). "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66–73.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Miyata, H., Nomura, H., and Houtgast, T. (**1991**). "Speech-intelligibility and subjective MTF under diotic and dichotic-listening conditions in reverberant sound fields," Acustica **73**, 200–207.

Schroeder, M. R. (**1981**). "Modulation transfer-functions: Definition and measurement," Acustica **49**, 179–182.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Viemeister, N. F. (**1979.**) "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364-1380.