Fast assessment of auditory spectral and temporal resolution

Arne Pelzer[1], Sébastien Santurette[2], Federica Bianchi[3], Torsten Dau[4]

[1] *Technical University of Denmark, 2800 Kongens Lyngby, Denmark, Email: arnepelzerprivat@gmail.com*

[2] *Technical University of Denmark, 2800 Kongens Lyngby, Denmark, Email: ses@elektro.dtu.dk*

[3] *Technical University of Denmark, 2800 Kongens Lyngby, Denmark, Email: fbia@elektro.dtu.dk*

[4] *Technical University of Denmark, 2800 Kongens Lyngby, Denmark, Email: tdau@elektro.dtu.dk*

Introduction

Although measures of spectral and temporal resolution may be of great importance for characterizing individual hearing impairment, these are time consuming and therefore not currently used in clinical practice. A recent study [1] suggested correlations between outcomes of a fast test ("FT-Test", [2, 3]) and traditional measures of frequency and temporal resolution. However, the clinical relevance of the test remains unclear. Here, the FT-Test procedure and stimuli were refined and this modified method, referred to as mFT-Test, was tested in normal-hearing (NH) and hearing-impaired (HI) listeners, with the aim to investigate whether it can provide a valid indicator of frequency selectivity and of the ability to take advantage of temporal masker fluctuations in a speech-in-noise task. Detection thresholds of a pulsed tone in a stationary threshold-equalizing noise masker were compared to detection thresholds in a noise containing either a spectral notch or temporal modulation. Masking release values were then compared to conventional measures of spectral and temporal resolution in the form of auditory filter bandwidths derived from a notched-noise experiment and speech reception thresholds in stationary versus modulated maskers. If the mFT-Test is a valid assessor of spectral and temporal resolution, masking release values in this test should be significantly correlated with the outcomes of the more conventional measures.

Method

Six NH listeners (aged 23 to 63 years, median 28) and eleven HI listeners (aged 52 to 80 years, median 70) participated in the study. All HI listeners showed air-bone gaps of at most 10 dB at all audiogram frequencies and had a maximum hearing loss of 65 dB HL at audiogram frequencies from 0.125 to 3 kHz. NH listeners had audiometric thresholds of at most 20 dB HL at all frequencies. All participants were native Danish speakers. Except for the pure tone audiometry, all measurements were conducted monaurally on the same ear. For both NH and HI listeners, the ear with the lower absolute threshold at 3 kHz was selected as the measurement ear. If there was no difference between left and right thresholds at 3 kHz, the ear with lower mean threshold was selected. Figure 1 shows the hearing thresholds of the ears used in this study.

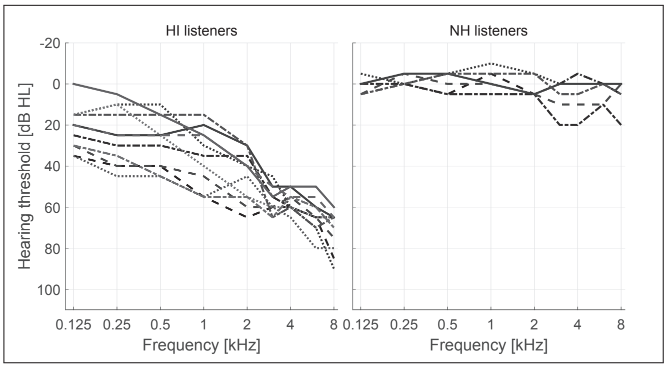

Figure 1: Pure tone audiograms of the 11 HI ears (**left panel**) and 6 NH ears (**right panel**) used in the study.

mFT-Test

A Békésy tracking procedure was used to estimate masked detection thresholds of pulsed tones in noise. In each run, tone pulses with fixed frequency were used as the target signal. A pulse was presented every 450 ms, yielding a pulse rate f_p of $2.\bar{2}$ Hz. Tone pulses had a duration of 20 ms with 5 ms rise/fall time and a 10 ms steady-state portion, resulting in an inter-pulse interval t_{IPI} of 430 ms. Threshold-equalizing noise (TEN, [4]) was used to mask the tone pulses. TEN is spectrally shaped such that it gives equal masked thresholds for pure tone signals at all frequencies between 250 Hz and 10 kHz for NH listeners. Its level is specified in dB/ERB. ERB refers to the mean equivalent rectangular bandwidth of the auditory filter at a chosen frequency for young NH listeners [5].

Three different noise maskers were used in the mFT-Test. As a reference for the other two maskers, a 6-ERB wide TEN centered at the target frequency was used. The three noise conditions derived from this were as follows:

1. The original 6-ERB wide TEN band was considered as the reference condition
2. A 6-ERB wide TEN band with a spectral notch of three ERBs width centered at the target frequency was used to assess spectral resolution
3. A 6-ERB wide TEN band with temporal amplitude modulation was used to assess temporal resolution

All signals were generated digitally in Matlab R2014a (The Mathworks). Spectral notches were inserted in the frequency domain by setting the corresponding frequency bins to zero. Amplitude modulation was achieved by multiplying the discrete time domain noise signal with the modulation function given by equation 1, where n is the discrete time index.

$$mod(n) = 1 + \cos(2\pi f_m n + \varphi_m) \quad (1)$$

The modulation frequency f_m was fixed at twice the pulse rate f_p of the signal, yielding $f_m = 4.\bar{4}$ Hz. The modulator phase φ_m was chosen such that every signal pulse was temporally centered in a modulation dip of the masker. This lead to the target signal being presented in every second modulation dip of the masker in the modulated condition 3 (see fig. 2, bottom panel). The specific modulation frequency was chosen because it is close to 4 Hz. Earlier studies suggest that most of the useful linguistic information in speech modulation is contained at modulation frequencies around 4 Hz [6]. The temporal relationship between signal pulses and masker is illustrated in fig. 2. The noise masker in different noise conditions is shown in the frequency domain in fig. 3. Both the introduction of a spectral notch and of temporal modulation into the masking noise remove a certain amount of masking power, thereby potentially lowering the masked detection threshold of the pulsed tones in comparison to the reference condition. Temporal and spectral resolution were expressed as the amount of release of masking (RoM), relative to the reference condition, as exhibited by the conditions with spectral notch or temporal modulation. The mFT-Test was carried out at signal frequencies of 500 Hz and 3 kHz. The masker level was set at 55 dB SPL/ERB when the absolute hearing threshold was 45 dB HL or less and at 10 dB SL/ERB relative to the absolute hearing threshold when it was 50 dB HL or more. Subjects received one training run for every combination of noise condition and signal frequency. Subsequently, three repetitions were measured for every noise condition in randomized order at both signal frequencies. The resulting threshold estimations were averaged within their respective categories.

DaHINT

Speech understanding in stationary and modulated speech-shaped noise was measured using the Danish Hearing In Noise Test, referred to as DaHINT [7]. For stationary and modulated noise, listeners were presented with lists of 20 Danish sentences each, masked by noise with a fixed level, and instructed to repeat the sentences as they understood them. The speech level was changed adaptively after each sentence depending on whether the subject was able to reproduce the last sentence correctly or not. For each sentence list, the subject's speech reception threshold in noise (SRT_N) was estimated as the signal-to-noise ratio at which 50% of the sentences were reproduced correctly. The procedure is described in more detail in [7]. The noise masker was synthesized by filtering white Gaussian noise with a 2048th order finite impulse response filter with a transfer function approximating the long term average spectrum of the DaHINT test speech material. This synthesized noise masker was used for the stationary noise condition of the DaHINT.

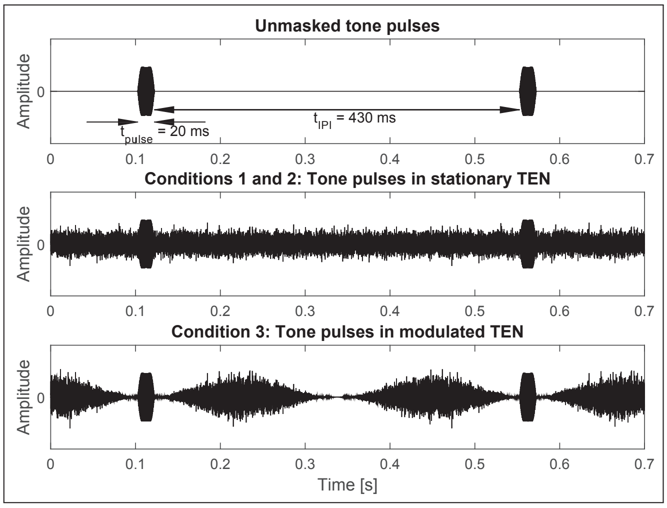

Figure 2: Temporal relationship of signal and masker of the mFT-Test. **Top panel**: Tone pulses without masker. **Center panel**: Tone pulses masked by stationary noise in reference condition (1) and spectral condition (2). **Bottom panel**: Tone pulses masked by modulated noise in temporal condition (3).

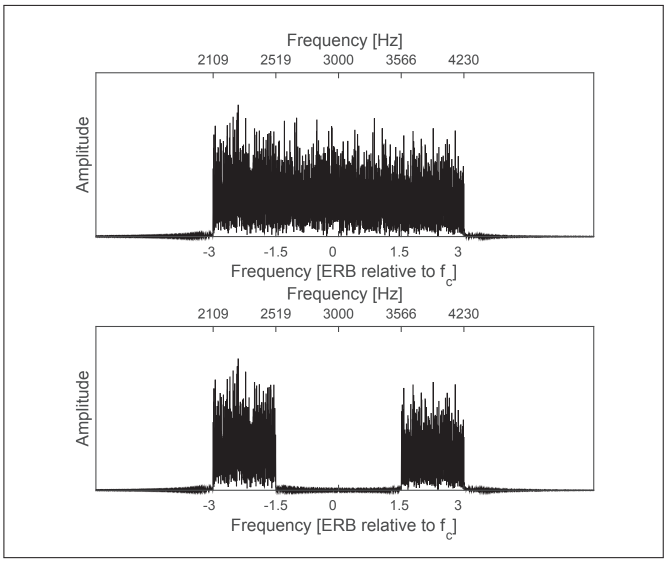

Figure 3: mFT-Test masker in the frequency domain. **Top Panel**: Spectrum of a 6 ERBs wide masking noise without notch centered at a signal frequency of 3 kHz as used in reference condition (1) and temporal condition (3). **Bottom panel**: Spectrum of a 6-ERB wide masking noise with a 3-ERB wide notch centered at a signal frequency of 3 kHz as used in spectral condition (2).

For the modulated noise condition, the noise masker was modulated in the same way as the noise in the temporal condition of the mFT-Test, i.e. with a modulation frequency $f_m = 4.\bar{4}$ Hz. However, the modulator phase φ_m was varied randomly in each trial to make sure that the positions of the modulation dips of the masking noise relative to the presented sentences were different in every trial. Temporal resolution was expressed as the RoM for the modulated noise condition relative to the condition with stationary noise. Obtained RoM values were compared with RoM values for the temporal condition of the mFT-Test measured at both 500 Hz and 3 kHz.

Notched-noise experiment

Frequency selectivity was measured by estimating the shape of the auditory filters centered at target frequencies of 500 Hz and 3 kHz using a notched-noise paradigm [8]. Subjects had to detect sine tones with a fixed level and a duration of 440 ms that were temporally centered in noise maskers of 550 ms duration. The masker level was varied adaptively based on subject response. Fixed-amplitude, random-phase noise was generated in the spectral domain. Spectral band limitation and notches were realized by setting the corresponding frequency bins to zero. The notch bandwidth was set relative to the target frequency f_0 as $\delta f/f_0$, where δf denotes the spectral distance between the inner masker edges and the target frequency f_0. Seven notch conditions were used, five symmetric ($\delta f/f_0$: 0, 0.1, 0.2, 0.3, 0.4) and two asymmetric ones ($\delta f/f_0$: 0.2|0.4, 0.4|0.2). The outer masker edges were fixed at $\pm$ 0.8 f_0. A three-alternative forced-choice (3AFC) weighted up-down method was used to track the 75% correct point on the psychometric function. Auditory filter shapes were fitted to the recorded data using the nonlinear least-squares method described in [9] and used to calculate the ERB of the filters. RoM values from the spectral condition of the mFT-Test were compared with the ERBs obtained from the notched-noise experiment.

Results

Figure 4 shows the mean spectral and temporal resolution of NH and HI listeners, expressed as RoM between the reference condition and either the spectral or temporal condition of the mFT-Test, at 500 Hz and 3 kHz. RoM values were larger for the NH subjects across frequencies and conditions. For NH listeners, spectral RoM was similar across frequencies, with RoM values of 18.3 dB at 500 Hz and 19.9 dB at 3 kHz. In the temporal condition, NH listeners could take better advantage of the masker modulation at 3 kHz, where the RoM was 15.3 dB, compared to 7.5 dB at 500 Hz. HI listeners showed smaller RoM across conditions and frequencies compared to the NH listeners. In both conditions, HI performance was better at 500 Hz than at 3 kHz, with spectral RoM dropping from 12.1 dB at 500 Hz to 2.4 dB at 3 kHz and temporal RoM dropping from 2.6 to -1.3 dB beween 500 Hz and 3 kHz.

Figure 5 shows scatter plots of the ERBs of all subjects versus the RoM from the spectral condition of the mFT-Test at target frequencies of 500 Hz and 3 kHz. The correlation between ERBs and spectral RoM at 500 Hz was significant when calculated for NH and HI listeners together ($r = -0.50$, $p < 0.05$), but not significant when calculated for the HI listeners alone. Despite this, linear regression fits calculated for the total group and the HI group show the same trend of decreasing spectral RoM with rising filter bandwidth. At 3 kHz target frequency, again the correlation for the data of NH and HI listeners together is significant ($r = -0.54$, $p < 0.05$), while that for the HI listeners alone is not. In this case, linear regression fits for the total group and the HI group are not in good agreement with each other.

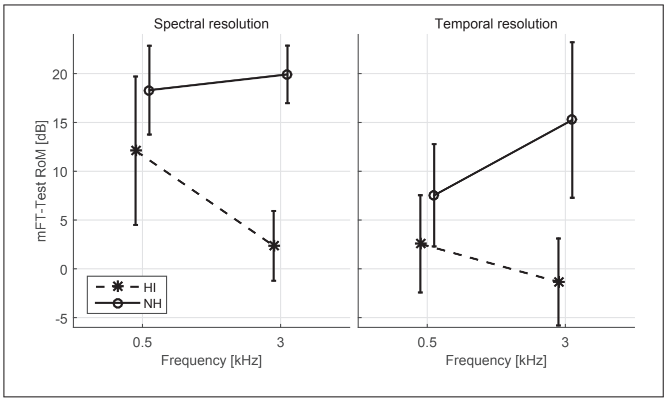

Figure 4: Mean spectral resolution (**left panel**) and temporal resolution (**right panel**) of NH and HI listeners, expressed as RoM between the reference condition and either the spectral or the temporal condition of the mFT-Test. Errorbars represent ± 1 standard deviation.

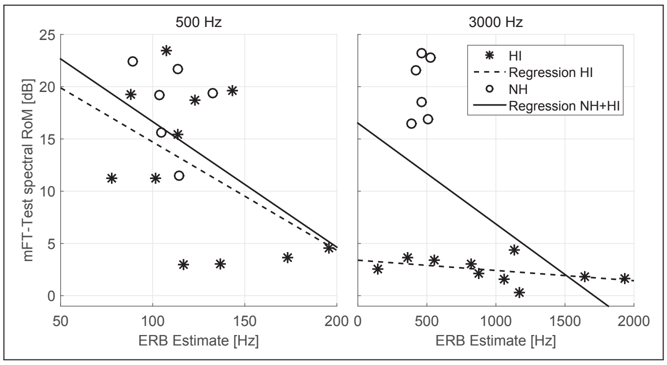

Figure 5: Scatter plots of ERBs from a notched-noise experiment versus spectral resolution from the mFT-Test at target frequencies of 500 Hz (**left panel**) and 3 kHz (**right panel**).

The relationship between the RoM observed between the stationary and modulated noise conditions of the DaHINT and temporal RoM from the mFT-Test is shown in fig. 6. The correlation of the DaHINT results with temporal RoM values at 500 Hz was highly significant for NH and HI listeners together ($r = 0.74$, $p < 0.01$) as well as for HI listeners alone ($r = 0.71$, $p < 0.01$), and linear regression fits for the total group and HI listeners alone show the same trend of increasing temporal RoM with increasing DaHINT RoM. Correlations of the same DaHINT data with the temporal RoM values at 3 kHz were highly significant when calculated for the total group ($r = 0.76$, $p < 0.01$) and not significant for the HI listeners alone. Linear regression fits for the two groups showed the same trend of increasing temporal RoM with increasing DaHINT RoM.

Discussion

As there was no significant correlation between the spectral RoM from the mFT-Test and the individual ERB estimates from the notched-noise experiment for the HI

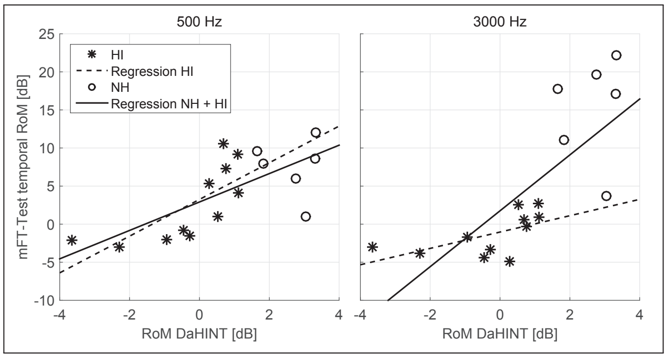

Figure 6: Scatter plots of RoM values from the DaHINT versus temporal resolution from the mFT-Test at target frequencies of 500 Hz (**left panel**) and 3 kHz (**right panel**).

listeners alone at either 500 Hz or 3 kHz, no conclusion about the validity of the mFT-Test as an assessor of spectral resolution could be drawn here. At 3 kHz target frequency, the spread of the RoM values of the mFT-Test was low for the HI listeners, with dynamic ranges of 4.1 dB for the spectral RoM values and 7.7 dB for the temporal RoM values. This low spread was in part caused by the fact that the mFT-Test masker level at 3 kHz target frequency was set at 10 dB SL/ERB relative to the absolute hearing threshold, thereby limiting the possible amount of RoM for the spectral and temporal condition relative to the reference condition. Additionally, ERB results from the notched-noise experiment showed some HI listeners with sharply tuned auditory filters at 3 kHz. This is surprising, considering that these subjects have high-frequency hearing loss, which is usually associated with a broadening of the respective auditory filters. No explanation for this observation could be found.

The correlation of the temporal RoM values from the mFT-Test and the RoM values from the DaHINT was highly significant at 500 Hz, both for the total group and the HI listeners alone. This finding suggests that the mFT-Test is suited to assess temporal resolution, expressed as the RoM in a speech-in-noise task for a modulated noise condition relative to a stationary one.

The mean duration of the mFT-Test per ear and frequency was 14 minutes for one training run and three measurement repetitions. The median standard deviation of the three measurement repetitions, averaged across listeners and conditions, was 1.13 dB. Because of this high stability, the duration could be further reduced by lowering the number of repetitions of the measurement runs to two or one, while keeping the training run to ensure subject familiarity with the task. In contrast, the duration of the notched-noise experiment to determine auditory filter bandwidths was in the order of 1.5 hours per ear and frequency, while the DaHINT measurement for one ear took approximately half an hour. Therefore, the mFT-Test can be considered a very fast test that is well-suited for clinical procedures from the perspective of time-expenditure.

Conclusion

The mFT-Test was found to be a valid assessor of temporal resolution. No conclusion about the validity of the mFT-Test as an assessor of spectral resolution could be drawn. For future use of the mFT-Test, care should be taken to use the same masker sensation level at all target frequencies. The masker sensation level should be sufficiently high to enable large RoM values between the spectral and temporal condition and the reference condition of the mFT-Test. It is possible that this would result in a larger spread of the spectral and temporal RoM values. An improved version of the mFT-Test taking this into account might better reflect spectral resolution. The mFT-Test was found to be significantly faster than other measures of spectral and temporal resolution, making it well-suited for use in a clinical setting.

References

[1] van Esch, T.; Dreschler, W.: Measuring spectral and temporal resolution simultaneously: a comparison between two tests. International Journal of Audiology (2011), 50(7):477–90.

[2] Larsby, B.; Arlinger, S.: A Method for Evaluating Temporal, Spectral and Combined Temporal-Spectral Resolution of Hearing. Scandinavian Audiology (1998), 27(1):3–12

[3] Larsby, B.; Arlinger, S.: Auditory temporal and spectral resolution in normal and impaired hearing. Journal of the American Academy of Audiology (1999), 10(4):198–210.

[4] Moore, B.; Huss, M.; Vickers, D.; Glasberg, D.; Alcantara, J.: Test for the Diagnosis of Dead Regions in the Cochlea. British Journal of Audiology (2000), 34(4):205–224

[5] Glasberg, B.; Moore, B.: Derivation of auditory filter shapes from notched-noise data. Hearing Research (1990), 47(1-2):103–138.

[6] Drullman, R.; Festen, J.; Plomp, R.: Effect of temporal envelope smearing on speech reception. The Journal of the Acoustical Society of America (1994), 95(February 1994):1053–1064.

[7] Nielsen, J.; Dau, T.: The Danish hearing in noise test. International Journal of Audiology (2011), 50:202–208.

[8] Patterson, R.; Nimmo-Smith, I.; Off-frequency listening and auditory-filter asymmetry. The Journal of the Acoustical Society of America (1980), 67(1):229.

[9] Strelcyk, O.; Dau, T.: Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. The Journal of the Acoustical Society of America (2009), 125(5):3328–45.