Assessing Differences in Music Perception Using Music-In-Noise Testing (MINT)

John Kyle Cooper and Shafaqat Rahman
Department of Biomedical Engineering, University of Rochester
jcoop32@ur.rochester.edu, srahm11@ur.rochester.edu

ABSTRACT

Recent research in music cognition suggests that individuals with musical training may experience less difficulty understanding speech in noise than individuals without musical training. In order to further validate this claim, a new assessment called the Music-In-Noise Test (MINT) was used to compare the pitch perception of normal hearing (NH) individuals with and without musical training. Our results show that NH individuals with no musical training can achieve a MINT accuracy of 75% at an SNR of -10.5 dB (n = 3) while NH individuals with musical training can achieve the same accuracy level at an SNR of -13.5 dB (n = 1). The results of this study further suggest that musical training may improve the pitch perception of NH individuals.

1. INTRODUCTION

The ability to isolate target sounds from background noise is a well sought after research topic in neuroscience and computer audition. Poor speech understanding in noise is a common difficulty reported amongst individuals with hearing impairments (Wong et al., 2013). However, recent literature in neuroscience suggests that people with musical training may experience less difficulty understanding speech in noise than people without musical training.

Cheng et al. (2018) assessed the effects of musical training in pediatric Mandarin-speaking cochlear implant users. Their results showed that musical training can significantly improve the music and speech perception of Mandarin-speaking cochlear implant users. Based on these findings, we are interested in knowing if musical training can sharpen a child's ability to detect subtle spectral and temporal differences in a target signal at varying levels of noise. In order to further validate these findings, we propose a modified version of the Hearing in Noise Test (HINT).

The Hearing In Noise Test (HINT) is a standardized test primarily developed by Nilsson, et al. (1994) that evaluates speech recognition in noise, and does so by assessing a patient's ability to accurately discern the contents of a target signal at different signal-to-noise ratio (SNR) levels. The level of speech is adjusted with respect to a fixed noise level, and thus the SNR is calculated in reference to a modulating speaker level. Traditionally, the test is an adaptive assessment in which the SNR is lowered for every correct response from the patient while the SNR is increased for every incorrect response. The test converges so that the patient obtains a 50% accuracy rate at a unique SNR level, which can then be used to compare scores across different patients and different experimental conditions (noisy and impaired). Another version of the HINT is also used (quickSIN) which relies on similar principles of testing except that the noise is modulated at a fixed speaker level.

The proposed assessment is a modified version of the HINT and has been coined the Music in Noise Test (MINT). This assessment is used to evaluate an individual's pitch perception through a two-alternative forced-choice (2AFC) task, in which the individual compares two paired melodies. We will use MINT to test the hypothesis that normal hearing (NH) individuals with musical training will have better pitch perception in noise than NH individuals without musical training.

2 METHODS

2.1 Dataset

Synthetic piano recordings and an isolated speech-like noise source were provided by Dr. Anne Luebke's lab at the University of Rochester. These piano recordings were modified to have changes in pitch, scale, and contour in their melodies (see Appendix A). Recordings were synthesized so that two melodies would be played with a fixed pause in between. Recordings ending with a '0' indicate the same melody being played between the pause. Recordings ending with '1' indicate an out of scale change, '2' a contour change, and '3' an interval change. Each melody only appears once in a list and each list contains two subsets of equally balanced recording types (12 same, 12 different). Within each list, the subset of 12 different recordings is balanced to have 4 out of scale changes, 4 contour changes, and 4 interval changes.

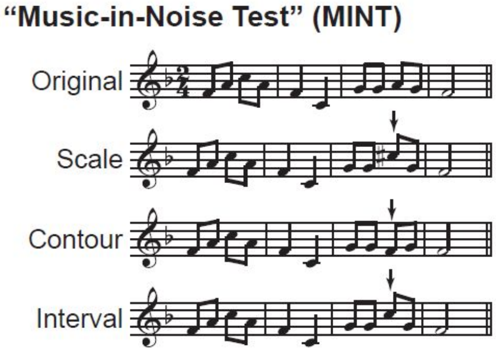


Figure 1. Different types of melodies in the MINT.

A total of 6 lists are used for testing but are adjusted to have different SNR values.

$$SNR\ (dB) = 20 * log(\frac{RMS\ SIGNAL}{RMS\ NOISE}) \qquad \text{Eq. 1}$$

We applied the fixed noise to each recording and modulated the target melody's level via MATLAB. The SNR levels for each recording in each list were calculated using equation 1. The first, second, third, fourth, and fifth lists contain recordings at 0, -6, -9, -12, and -15 dB SNR. The final list contains recordings at zero noise (quiet condition).

2.2 YIN Algorithm: Validating Pitch Differences Amongst Recording Types

Currently, there are few datasets dedicated to the MINT. For future use of MINT, we would like to utilize datasets constructed by other researchers. However, in order to maintain consistency between datasets, we need to validate these datasets before use in MINT. Therefore, a workflow was created in this study to validate which files in the provided MINT dataset are designated as "same", "out of scale change", "contour change", "interval change", The main component of this workflow was the YIN pitch estimation algorithm (Cheveigné & Kawahara, 2002), which was used to generate pitch estimates for all recordings. Using the pitch estimates for each recording, we detected the differences in pitch between the two music segments within each recording to determine the type (shown in Figure 2).

2.3 Validation through Onset Detection

Furthermore, the type of each recording in every dataset is not the only attribute that needs to be validated. The onset timing of every note in each recording must be consistent between the paired melodies. Onset timing was evaluated to ensure that the change in pitch was the only attribute being tested between each melody within each recording. In order to compare onset intervals between melodies, we calculated the root-mean-square (RMS) of each melody in each recording using equation 2 to see the energy at each note played.

$$x_{RMS} = \sqrt{\frac{1}{n}(x_1{}^2 + x_2{}^2 + ... + x_n{}^2)} \qquad \text{Eq. 2}$$

Where x is a music segment array and n is the number of elements in the music segment array.

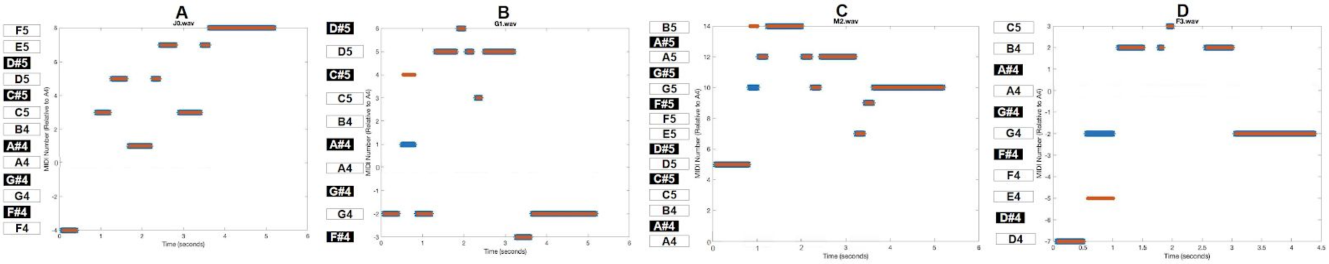


Figure 2. plots of the pitch estimates for MINT audio files in which A) the two music segments are the same B) the second music segment has one note that is an out of scale change C) a contour change D) an interval change.

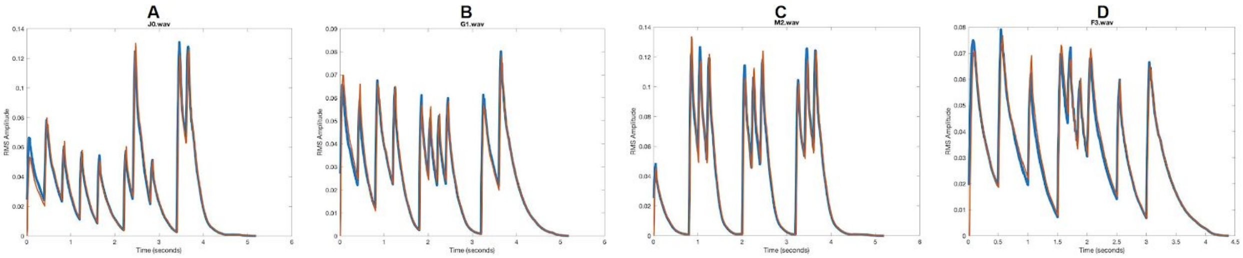


Figure 3. RMS plots of MINT audio files in which A) the two music segments are the same B) the second music segment has one note that is an out of scale change C) a contour change D) an interval change.

Our comparisons (shown in Figure 3) revealed that the onset intervals did not change between music segments. We then claimed that pitch was the only attribute being tested between each music segment within each audio file.

2.4 Implementation of MINT

In order to carry out the MINT, we designed a GUI in MATLAB to play the recordings in each list and store the user's input (shown in Figure 4). The input is strictly binary; if the user believes during the pairwise comparison that the melodies are the same, they click the green 'same' button (user response will be stored as a '1'). If they believe the melodies sound different, they click the red 'different button (user response will be stored as a '0'). The user does not have any knowledge of any of the changes we made to scale, contour, and interval. The first part of the test involves the participant to go through a practice run with recordings in quiet (i.e. without noise). The second part of the test involves the participant to go through all 6 lists at varying SNR with no breaks in between and takes about 15 minutes to complete. Results are exported to Microsoft Excel for data analysis.

During the test, if the melody sounds the same, click the green 'same' button.
If the melody sounds diffferent, click the red 'different' button.

Figure 4. The MINT GUI interface for carrying out the music-in-noise perceptual test. The application was created in MATLAB 2018b.

2.5 Participants

For this study, we recruited 3 NH subjects with no musical training (mean age = 25 years) and 1 NH subject with 10 years of musical training (age = 22 years).

3. RESULTS

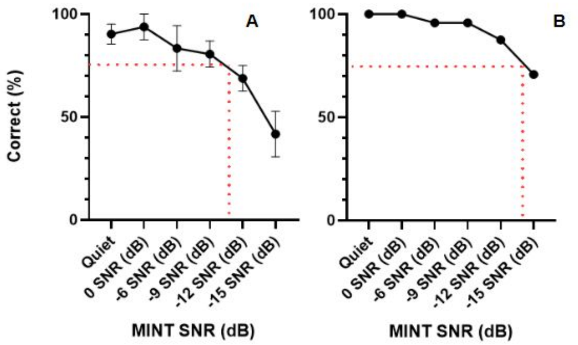


Figure 5. MINT plots for participants with A) normal hearing and no musical training (n = 3) and B) normal hearing with formal musical training (n = 1)

The MINT examination was carried out on participants with normal hearing and no musical training (n = 3) and with formal musical training (n = 1). Based on the collected MINT data, the point at which a person with normal hearing and no musical training can achieve a 75% accuracy is at an SNR of -10.5 dB (Figure 5). SNRs lower than this value lead to a diminished accuracy in the pooled participants. For our participant with formal musical training, the SNR at which he could obtain 75% accuracy occurs at -13.5 dB.

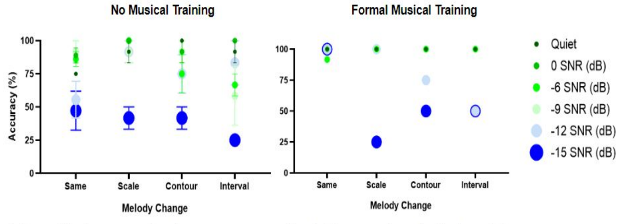


Figure 6. Accuracy plots in respect to the type of melodic comparison for both participants with no musical training (left) and with formal musical training (right)

In Figure 6, between the range of conditions from quiet to -3 SNR dB, the accuracy at which participants could detect a same comparison or a different comparison with a scale, contour, or interval change is nearly around 80-100% regardless of having formal musical training. However, participants with no musical training showed a decreased performance at accurately determining the pairwise comparisons with increasing noise level. This differs from our participants with musical training, as he was able to detect all the same pairwise comparisons even up to -15 SNR dB, and showed difficulty in determining the other melodic types past -12 SNR dB.

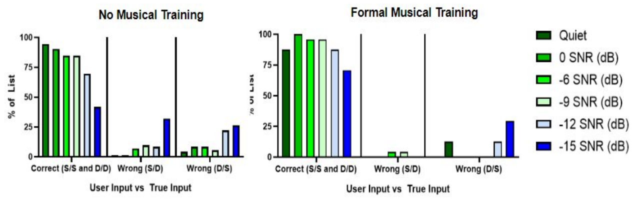


Figure 7. Percent distribution of user input compared to true value for conditions in which the user gets a correct (S/S, D/D) or an incorrect (S/D, D/S) response for participants with and without musical training.

We did not see any differences between the participants with no musical training in regards to noise biasing their ability to make a correct decision (Figure 7). For both situations with increasing noise, this cohort's likelihood for deciding a comparison is the same when it is different (S/D) or vice versa (D/S) is balanced. We cannot make the same assumption for people with musical training, again due to our low sample size.

4. DISCUSSION

Based on our results, we believe that musical training can enhance a person's ability to detect subtle changes in a target signal in a noisy environment. Despite our low sample size for both cohorts, Figure 5 suggests that people with musical training can have enhanced perceptual discernment, at least in the context of the study's parameters. The participant with musical training had a 75% accuracy point at -13.5 dB compared to participants with no formal training at -10.5 dB. This is a 3 dB difference in increased sensitivity, suggesting musical training may significantly improve the perceptual sensitivity of a participant at detecting subtle melodic changes.

However, no statistics were done to test for statistical significance due to low sample sizes in both cohorts. Another interesting point of the study is found in Figure 6, where a clear trend can be seen when the melodic comparisons are altered to have an out of scale, contour, or interval change. Participants do not seem to experience much difficulty in assessing out of scale changes up until -15 SNR dB, but do experience difficulty with contour or interval changes at increasing SNR (distributions for these types are much more varied).

The study is currently limited in our ability to determine the 75% accuracy rate for people with hearing impairments. Based on these preliminary results, we believe a person without musical training and experiencing hearing difficulties may have a 75% accuracy SNR much lower than -10.5 dB, and that with musical training, that point will shift to the right.

In addition, we are currently lacking in our ability to determine a neural threshold at which people can detect differences in a target signal. The study was designed to provide for clear spectral differences in melody, contour, or interval, as only one note is being adjusted in each melodic comparison. Through the YIN algorithm, we should be able to quantify and determine a threshold point in future iterations of the study.

5. CONCLUSIONS

In this study we used MINT to determine the thresholds that humans can discern differences in paired melodies. Currently, our experiment is lacking in terms of data from participants with musical training, since we were only able to recruit 1 subject with this criterion. Using the YIN algorithm, we generated pitch contours that can visibly showcase the pitch differences in musical melodies that are out of scale, contour, or interval. We also calculated the RMS of each melody within each recording to detect onsets, and show that the onsets do not differ across the melody types. Thus, the participant's decision on whether paired melodies are the same or different must come from the pitch and not from the beat. Based on our preliminary MINT data, we find that NH individuals with musical training can achieve a MINT accuracy of 75% at a lower SNR level (i.e. higher noise level) than NH individuals with no musical training.

5. FUTURE WORK

Further MINT assessments are required to concretely assess the 75% accuracy SNR level for NH participants with and without musical training. We would also like to use MINT to assess how musical training influences the 75% accuracy SNR level for hearing-impaired participants. In addition, we would like to make modifications to the MINT so that it is more adaptive and converges onto an SNR at 75% accuracy during the examination, as the current test requires data analysis to obtain the 75% threshold. The information from this assessment can further elucidate the benefits of using music as a therapy to improve our brain's ability to source separate stimuli from noisy environments. Furthermore, we would like to investigate in future studies if people with musical training can also better detect differences in general speech in a noisy environment.

6. REFERENCES

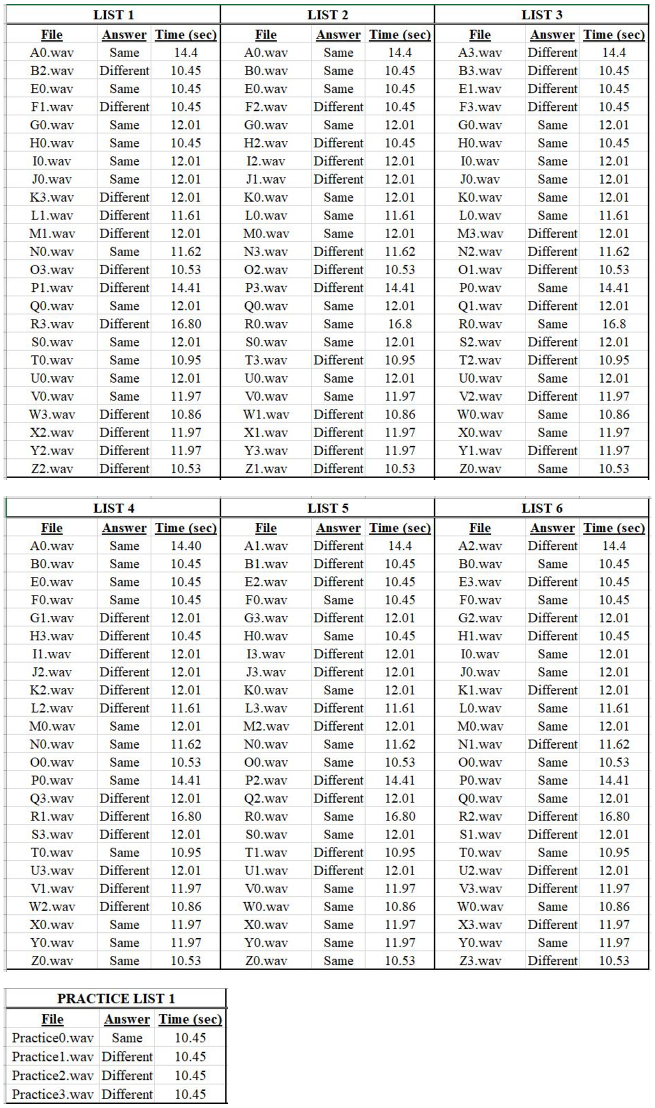
[1] Wong, L. L. N., Hickson, L., & McPherson, B. (2003). Hearing aid satisfaction: what does research from the past 20 years say? *Trends in amplification, 7*(4), 117-161. doi:10.1177/108471380300700402

[2] Cheng, X., Liu, Y., Shu, Y., Tao, D. D., Wang, B., Yuan, Y., . . . Chen, B. (2018). Music Training Can Improve Music and Speech Perception in Pediatric Mandarin-Speaking Cochlear Implant Users. *Trends Hear, 22*, 2331216518759214. doi:10.1177/2331216518759214

[3] Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet

and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085-1099. doi:10.1121/1.408469
[4] Cheveigné, A., & Kawahara, H. (2002). YIN, A fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America, 111*, 1917-1930. doi:10.1121/1.1458024

7. APPENDIX A



LIST 1			LIST 2			LIST 3		
File	**Answer**	**Time (sec)**	**File**	**Answer**	**Time (sec)**	**File**	**Answer**	**Time (sec)**
A0.wav	Same	14.4	A0.wav	Same	14.4	A3.wav	Different	14.4
B2.wav	Different	10.45	B0.wav	Same	10.45	B3.wav	Different	10.45
E0.wav	Same	10.45	E0.wav	Same	10.45	E1.wav	Different	10.45
F1.wav	Different	10.45	F2.wav	Different	10.45	F3.wav	Different	10.45
G0.wav	Same	12.01	G0.wav	Same	12.01	G0.wav	Same	12.01
H0.wav	Same	10.45	H2.wav	Different	10.45	H0.wav	Same	10.45
I0.wav	Same	12.01	I2.wav	Different	12.01	I0.wav	Same	12.01
J0.wav	Same	12.01	J1.wav	Different	12.01	J0.wav	Same	12.01
K3.wav	Different	12.01	K0.wav	Same	12.01	K0.wav	Same	12.01
L1.wav	Different	11.61	L0.wav	Same	11.61	L0.wav	Same	11.61
M1.wav	Different	12.01	M0.wav	Same	12.01	M3.wav	Different	12.01
N0.wav	Same	11.62	N3.wav	Different	11.62	N2.wav	Different	11.62
O3.wav	Different	10.53	O2.wav	Different	10.53	O1.wav	Different	10.53
P1.wav	Different	14.41	P3.wav	Different	14.41	P0.wav	Same	14.41
Q0.wav	Same	12.01	Q0.wav	Same	12.01	Q1.wav	Different	12.01
R3.wav	Different	16.80	R0.wav	Same	16.8	R0.wav	Same	16.8
S0.wav	Same	12.01	S0.wav	Same	12.01	S2.wav	Different	12.01
T0.wav	Same	10.95	T3.wav	Different	10.95	T2.wav	Different	10.95
U0.wav	Same	12.01	U0.wav	Same	12.01	U0.wav	Same	12.01
V0.wav	Same	11.97	V0.wav	Same	11.97	V2.wav	Different	11.97
W3.wav	Different	10.86	W1.wav	Different	10.86	W0.wav	Same	10.86
X2.wav	Different	11.97	X1.wav	Different	11.97	X0.wav	Same	11.97
Y2.wav	Different	11.97	Y3.wav	Different	11.97	Y1.wav	Different	11.97
Z2.wav	Different	10.53	Z1.wav	Different	10.53	Z0.wav	Same	10.53

LIST 4			LIST 5			LIST 6		
File	**Answer**	**Time (sec)**	**File**	**Answer**	**Time (sec)**	**File**	**Answer**	**Time (sec)**
A0.wav	Same	14.40	A1.wav	Different	14.4	A2.wav	Different	14.4
B0.wav	Same	10.45	B1.wav	Different	10.45	B0.wav	Same	10.45
E0.wav	Same	10.45	E2.wav	Different	10.45	E3.wav	Different	10.45
F0.wav	Same	10.45	F0.wav	Same	10.45	F0.wav	Same	10.45
G1.wav	Different	12.01	G3.wav	Different	12.01	G2.wav	Different	12.01
H3.wav	Different	10.45	H0.wav	Same	10.45	H1.wav	Different	10.45
I1.wav	Different	12.01	I3.wav	Different	12.01	I0.wav	Same	12.01
J2.wav	Different	12.01	J3.wav	Different	12.01	J0.wav	Same	12.01
K2.wav	Different	12.01	K0.wav	Same	12.01	K1.wav	Different	12.01
L2.wav	Different	11.61	L3.wav	Different	11.61	L0.wav	Same	11.61
M0.wav	Same	12.01	M2.wav	Different	12.01	M0.wav	Same	12.01
N0.wav	Same	11.62	N0.wav	Same	11.62	N1.wav	Different	11.62
O0.wav	Same	10.53	O0.wav	Same	10.53	O0.wav	Same	10.53
P0.wav	Same	14.41	P2.wav	Different	14.41	P0.wav	Same	14.41
Q3.wav	Different	12.01	Q2.wav	Different	12.01	Q0.wav	Same	12.01
R1.wav	Different	16.80	R0.wav	Same	16.80	R2.wav	Different	16.80
S3.wav	Different	12.01	S0.wav	Same	12.01	S1.wav	Different	12.01
T0.wav	Same	10.95	T1.wav	Different	10.95	T0.wav	Same	10.95
U3.wav	Different	12.01	U1.wav	Different	12.01	U2.wav	Different	12.01
V1.wav	Different	11.97	V0.wav	Same	11.97	V3.wav	Different	11.97
W2.wav	Different	10.86	W0.wav	Same	10.86	W0.wav	Same	10.86
X0.wav	Same	11.97	X0.wav	Same	11.97	X3.wav	Different	11.97
Y0.wav	Same	11.97	Y0.wav	Same	11.97	Y0.wav	Same	11.97
Z0.wav	Same	10.53	Z0.wav	Same	10.53	Z3.wav	Different	10.53

PRACTICE LIST 1		
File	**Answer**	**Time (sec)**
Practice0.wav	Same	10.45
Practice1.wav	Different	10.45
Practice2.wav	Different	10.45
Practice3.wav	Different	10.45

Figure A. Lists of designated audio files with two melodies separated by a pause. Audio files are labeled with respect to the type of pairwise comparison (same, out of scale, contour, interval). Lists 1,2,3,4,5 and 6 are assigned to have audio files with SNRs of 0, -6, -9, -12, and -15 dB respectively. Practice List 1 contains only four audio files played at the beginning of the experiment.