Duration Thresholds for Identifying Different Sound Types

Durga Lal Budathoki[1], Jürgen Tchorz[2], Greg O'Beirne[1]

[1]University of Canterbury, Department of Communication Disorders, Christchurch, New Zealand

[2]Technische Hochschule Lübeck

Schlüsselwörter: Duration thresholds, sound classification, hearing loss screening

Introduction

It is well known that people with sensorineural hearing loss have difficulties with basic auditory processing abilities such as temporal processing, frequency selectivity, frequency discrimination, as well as difficulties with complex tasks such as speech perception in noise, music perception, and environmental sound awareness. Identifying and segregating short snippets of sound, and merging them into auditory streams is a prerequisite to accomplish these more complex tasks.

There have been multiple studies that have tested the time required to identify, detect, or classify short sounds using various methods for various purposes. The studies have either used a single sound class for identification of sound, or multiple sound classes for classification of sound types. Ballas et al. (1993) tested the time required to correctly identify environmental sounds which it defined as the sounds or noises from non-living things, excluding musical instrument. They found that the shortest time taken to correctly identify was for the sound of ringing telephone at 1253 ms. The longest time was for the sound of an electronic lock at 6823 ms. The reaction time required to initiate and perform movement for button press, after the mental identification of sound, was added within the identification time. Therefore, the identification time also includes the reaction time of the motor processing and motor actions for hand movements which varies across individuals. Thiesen et al. (2016) reported data on duration thresholds for classification of music genre. Individuals could classify music sounds to their correct genre even at a short duration of 250 ms. It has also been shown that the participants were more accurate at classifying stimuli when only the instrumental part of a song was presented, compared to stimuli where both vocal and instruments were presented (Gjerdigen & Perrot, 2008). Individuals without hearing problems could discriminate the emotional nature of classical music at very short duration ranging from 0.5 to 7.5 seconds (Peretz et al., 1998). Moradi et al. (2013) used gated stimuli in which they increased the duration of stimuli by 16.67 ms until the individuals correctly identified the consonants. The consonants were presented between two identical vowels. For example, if the stimuli was "aVa", participants had to correctly identify the consonant 'V'. The duration required to identify the consonants in quiet was significantly shorter that required in noise. However, in the beginning of the stimuli there was always a vowel sound, therefore the true duration of exposure to the consonant would be less than the actual threshold for their detection. Ogg et al. (2017) found that the participants were able to classify music, speech, and environmental sounds to their groups even when the sounds they heard were just 25 ms long. The sounds used were tested using gated go/no-go paradigm, starting with 12.5 ms duration and doubled in duration for each "gate". The go/no-go paradigm requires the participants to respond only when the target sound is played and not respond when other sounds are played. The minimum duration of the stimuli used was 12.5 ms, and the authors did not test the ability to classify sounds below 12.5 ms. Studies that use gated procedure only estimate the time range between which the sounds are identified, which means that there is always a risk of overestimating the identification thresholds. For example, in the study of Ogg et al. (2017), if a time threshold for identification was 15 ms, the participant does not respond at gated stimuli of 12.5 ms but they do respond at other stimuli at 25 ms. However, the true threshold is in between those two gated stimuli.

In a pilot study by Obert and Tchorz (2018), normal-hearing participants were required to listen to short sounds in a quiet environment and classify them into 4 categories: speech, noise, animal, and music, using a four alternative forced choice method (4AFC) with adaptive stimulus length. The duration thresholds for correct classification were class-dependent and between 21 - 45 ms. However, the minimum step size of 10 ms might have hindered precise threshold identification.

The aim of this study was to reliably measure the minimum duration of sounds needed to classify four major categories of sounds using a criterion-free 4AFC method in both normal hearing and hearing impaired listeners. In addition, the correlations between duration threshold, pure-tone thresholds and speech reception thresholds in noise were examined.

Material and Method

In total, 44 adult subjects participated in the trial. 16 of them had normal hearing and were between 21 and 63 years old (median: 34). 28 subjects (23-89 years, median: 69) had a sensorineural, symmetrical hearing loss. The pure-tone average (0.5 - 3 kHz) was between 18 and 60 dB HL (median: 44 dB). Each subject underwent pure tone threshold testing, and the University of Canterbury Auditory-Visual Matrix Sentence Test in auditory-alone mode (Trounson, 2012) to determine the speech reception threshold in noise.

The duration thresholds to identify different sound classes were measured using an adaptive 4AFC method with 4 simultaneous staircases and blockwise randomisation. After presentation of each sound, the subjects chose one of the four sound classes (speech, music, noise, animal) using a touch screen. After each trial, visual feedback was given on whether the answer was correct. An adaptive weighted up/down staircase method (Kärnbach, 1991) was applied to converge towards the threshold in the middle between chance (25%) and 100% correct. In case of a correct answer, the duration of the next example of this sound class was reduced by factor 0.88, and increased by factor 1.21 in case of a wrong answer.

The sound examples for each class were taken from various sources to ensure that the sound quality itself is no cue for classification. There were 200 sound files in total, 50 each for the four sound classes. Forty sounds from each sound class were used as real trials that counted towards calculation of time classification thresholds (averaging reversals of trial 21-40 for each sound class). Ten sound files from each class were used as decoys, which were not counted in the calculation for thresholds. The decoys were played once every fifth trial with the same duration as that of the class with the longest running duration. The decoys were introduced to prevent the duration of a sound from being used as additional cue. Single measurements were excluded from further analysis if there were more than 80% correct answers in the last 20 trials. After 2-3 weeks, the subjects were invited for a retest.

Results

Figure 1 shows the duration thresholds to identify different sound classes in normal hearing and hearing impaired subjects, based on the average values of test and retest, if available.

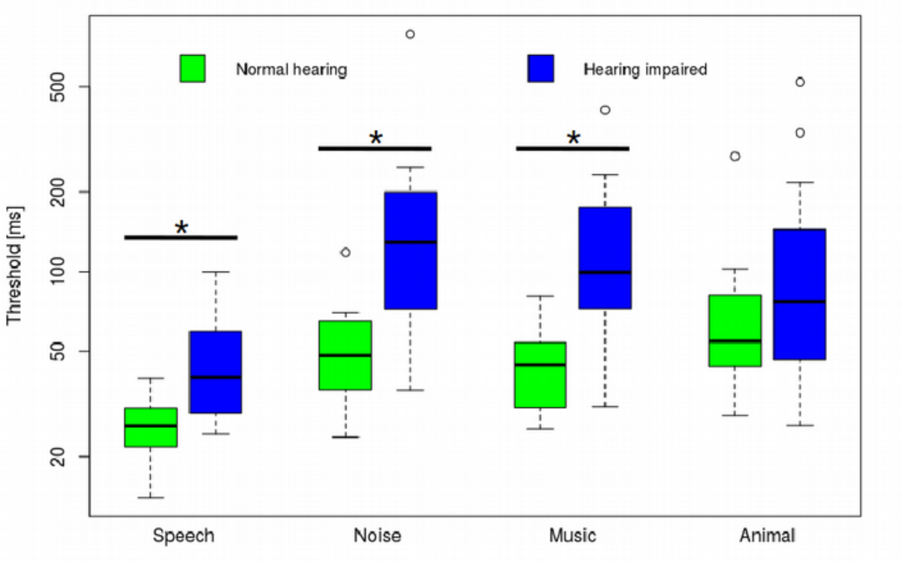

Figure 1: Duration thresholds for four different sound classes.

For speech, noise and music, there were significant differences between normal-hearing and hearing impaired subjects. This was not the case for animal sounds. Figure 2 shows the correlation between the hearing loss (pure-tone average, PTA), and the duration thresholds. The correlation is significant for speech, noise and music (Spearman rank correlation, $p < 0.05$), but not for music. Also, there is a significant correlation between the speech reception threshold (SRT) in noise and the duration thresholds for speech, noise and music (Spearman rank correlation, $p < 0.05$), but again not for music. A simple combined measurement value is the sum of all four duration thresholds for each subject. The correlation matrix for this combined value and SRT, PTA and high-frequency PTA (4-8 kHz) is given in Table 1.

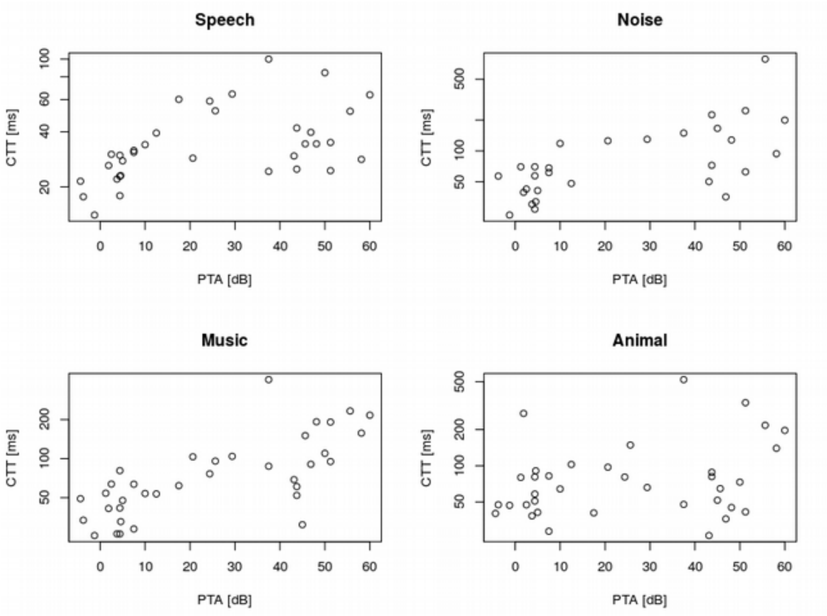

Figure 2: Correlations between the pure-tone average and duration thresholds for all four sound classes.

Table 1: Correlation matrix

	PTA	HF-PTA	SRT	Sum of Duration Thresholds
PTA	1	0.90	0.76	0.76
HF-PTA	0.90	1	0.85	0.73
SRT	0.76	0.85	1	0.68
Sum of D.T	0.76	0.73	0.68	1

The hearing loss (PTA) correlates to the same extend with (a) speech perception in noise and (b) the sum of all four duration thresholds. If the duration threshold for animal noise (which does not correlate to PTA or SRT) is excluded from the combined sum, the correlation coefficient between the sum of the remaining three duration thresholds and the PTA rises to 0.84.

The test-retest agreement can be examined using Bland-Altman plots (Figure 3, normal hearing subjects only). It can be seen that there is no systematic duration threshold difference between test and retest in speech and noise (the dashed line in the middle which indicates the mean difference between test and retest is close to 0 ms). For music and animal sounds, however, there is a trend towards lower duration thresholds in the retest, which might indicate a learning effect. The 95% confidence interval of the difference between test and retest in normal hearing listeners can be used as a threshold for detecting hearing loss. For music, for example, the 95% confidence interval is around 30 ms. From Figure 1 it can be seen that actually about ¾ of all hearing impaired subjects had a duration threshold for music which was more than 30 ms larger than the normal-hearing median (45 ms).

Conclusion

The duration thresholds needed for classification of different sound types can be measured using an adaptive 4AFC interleaved staircase procedure. In normal hearing listeners, these thresholds were found to be between about 20 and 80 ms, depending on the sound type. Duration thresholds for hearing impaired listeners were significantly longer in all sound types except animal, and they were also correlated to speech perception in noise. It turned out that the duration of an item might be used as additional cue for classification of the sound type. The presentation of decoy items was introduced to minimize this effect. A potential practical application of measuring duration thresholds might be a screening test for hearing loss which does not have tight demands on calibration (e.g., via smartphone apps). For such an application, however, it will be important to find sound snippets with same average duration thresholds both within and across sound classes.

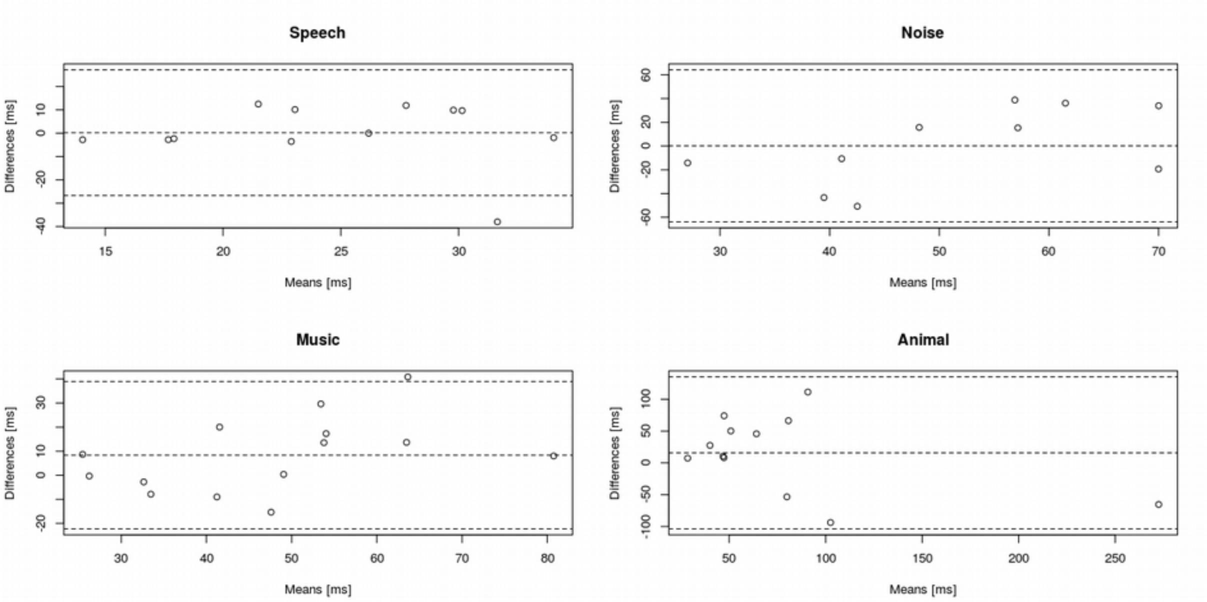

Figure 3: Bland–Altman plots with the mean (dashed line in the middle) and the 95% confidence interval of the difference between the two testing occasions.

Literature

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance, 9(2), 250-267.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The Rapid Recognition of MusicGenres. Journal of New Music Research, 37(2), 93-100.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. Percept & Psychophysiology, 49, 227-229.

Moradi, S., Lidestam, B., Rönnberg, J., (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. Frontiers in Psychology, 4(359), 359.

Obert, S., Tchorz, J., (2018). Geräuschklassifikation. Unpublished student project report, Lübeck University of Applied Science, Lübeck, Germany.

Ogg, M., Slevc, L. R., & Idsardi, W. J. (2017). The time course of sound category identification: Insights from acoustic features. The Journal of the Acoustical Society of America, 142(6), 3459-3473.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. Cognition, 68, 111–141.

Thiesen, F, C., Kopiez, R., Reuter, C., zedik-Eysenberg, I., & Schlemmer, K., (2016). In the Blink of an Ear: A Critical Review of Very Short Musical Elements (Plinks), presented at ICMPC14, July 5-9, 2016, San Francisco

Trounson, R. H. (2012). Development of the UC Auditoryvisual Matrix Sentence Test. Maud Thesis. Department of Communication Disorders. The University of Canterbury.