The Effects of Background Noise and Personal MP3 Player Volume on the Audibility to Bystanders

Brian M. Kreisman, PhD[1]
Kate M. Longenbach, AuD[2]
Melina A. Blasi, AuD[3]
Brian J. Fligor, ScD[4]
Andrew B. John, PhD[5]

[1] Calvin University, Department of Speech Pathology and Audiology, Grand Rapids, MI
[2] Specialty Physician Associates, Bethlehem, PA
[3] Ear, Nose & Throat Associates, Baltimore, MD
[4] Lantos Technologies, Woburn, MA
[5] University of Oklahoma Health Sciences Center, Department of Communication Sciences and Disorders, Oklahoma City, OK

Corresponding Author:
Brian M. Kreisman
Department of Speech Pathology and Audiology
Calvin University
3201 Burton SE
Grand Rapids, MI 49546
Phone: (616) 526-7306
Email: bmk24@calvin.edu

Conflict of Interest: The authors do not report any conflicts of interest.
Funding: Funding for the gift cards for participants was provided by Boston Children's Hospital.

ABSTRACT

Purpose: **To determine whether overhearing someone's iPod is a valid way to tell whether the user was listening at hazardous listening levels.** *Method:* **An iPod Touch with standard earbuds was placed on a KEMAR mannequin in a sound booth. The output was recorded and measured from 28 inches away to simulate a bystander. Recordings of five songs at nine iPod volume levels were mixed with seven background noise conditions: quiet, pink noise (45, 60, and 75 dBA) and real-world noise (45 dBA speech babble; 60 dBA restaurant; and 75 dBA airplane). Participants were seated in the center of an eight-speaker array. The song was played at 0-degrees azimuth with the background noise presented from seven speakers (45, 90, 135, 180, 225, 270, and 315 degrees azimuth). Participants, 50 young, normal-hearing adults, indicated if they heard the music during each trial.** *Results:* **Participants were likely to overhear someone's iPod in quiet; however, the probability that the song was at a hazardous listening level was low. As background noise levels increased, participants' ability to overhear iPods decreased. Positive predictive value increased as real-world noise increased; however, the pink background noise did not maintain this trend.** *Conclusion:* **Overhearing someone's iPod does not necessarily indicate they are listening at a hazardous level, and not overhearing someone's iPod does not necessarily indicate they are listening at a safe volume level. Thus, being able to overhear music emanating from someone's earphones is not a good indicator that iPod user is damaging his or her hearing.**

Key Words: **iPod, personal audio systems, MP3 player, listening levels, recreational noise–induced hearing loss, noise exposure**

INTRODUCTION

Approximately 48 million Americans experience some degree of hearing loss (National Institute on Deafness and Other Communication Disorders [NIDCD]; 2017). While in many instances the cause of hearing loss cannot be prevented, a large percentage of Americans suffer from noise-induced hearing loss, a type of hearing loss that is nearly 100% avoidable (NIDCD, 2017).

Recreational noise can be generated by numerous sources. Recent media attention has been focused on personal listening devices as a potential source for noise-induced hearing loss. Although such devices have been around for many years, their popularity has increased with the release of MP3 devices such as Apple's iPod in 2001 and smart phones. With the ubiquity of iPods, pocket MP3 players, and MP3 integrated devices such as cell phones, Americans may be running a higher risk of potentially exposing themselves to harmful doses of noise on a daily basis. The National Institute for Occupational Health and Safety [NIOSH] (2018) defines hazardous levels at 85 dBA for an 8-hour time-weighted average and uses a 3-dB exchange rate (for every increase of 3-dB, the recommended exposure time is halved).

Recently, the World Health Organization (WHO) and International Telecommunications Union (ITU) published *Recommendation ITU-T H.870* (August, 2018), which is an international standard for MP3 player and earphone manufacturers (Rec. ITU-T H.870 (08/2018), 2018; Safe listening devices and systems: a WHO-ITU standard, 2019). This standard was developed with the input of clinicians, manufacturers, governments, and health communication professionals to regulate exposure to loud sounds through "personal audio systems" and to mitigate the hearing loss risk associated with their use. This standard recommends all MP3 players include noise dosimetry as part of the operating system, controls for volume limiting and parental notification, and personalized messages and cues to action. The adoption of this global standard by WHO and ITU suggests an international consensus that MP3 players pose a serious and specific threat to hearing, and technological interventions are required to mitigate the threat.

Personal Listening Devices

Numerous studies have examined the noise exposure and potential damage to hearing that may be caused by personal listening devices (e.g., Bradley & Fortnum, 1987; Catalano & Levin, 1985; Clark, 1990; Fligor & Cox, 2004; Keith, Michaud, & Chiu, 2008; Danhauer et al., 2009; Feder, Marro, Keith, & Michaud, 2013; Gopal, Mills, Phillips, & Nandy, 2018; Hellstrom, Axelsson, & Costa, 1998; Hodgetts, Riegler, & Szarko, 2007; Hoover & Krishnamurti, 2010; Kreisman, 2014; Portnuff, Fligor, & Arehart, 2013; Pugsley, Stuart, Kalinowski & Armson, 1993; Turunen-Rise, Flottorp & Tvete, 1991; Wong, Van Hasselt, Tang, & Yiu, 1990). In one such study, Gopal and colleagues (2018) examined temporary threshold shifts in participants with normal hearing after listening to 30 minutes of music through a personal audio system (iPod Touch). Testing included pure-tone thresholds from .5 to 12.5 kHz and distortion product otoacoustic emissions (DPOAEs), both pre- and post-exposure. Groups were assigned an iPod volume level of either 0% (no music), 50%, 75% or 100%. Results suggested a temporary threshold shift for the group that listened to the music at 100% volume for pure-tone octaves and interoctaves from 2-8 kHz. In addition, that group experienced a decrease in DPOAEs at 2 and 2.822 kHz when compared to the other groups. The authors concluded that, for a 30-minute listening period, setting the volume at a level of 75% or lower would be considered safe based on the conditions of the study, while listening at 100% would expose the listener to a potentially dangerous noise dose.

Only one investigation of the use of MP3 players in adolescents was found. Twardella and colleagues (2017) surveyed 2143 ninth-graders on their use of MP3 players, including listening duration and volume level, and compared their results to available hearing tests. Their results suggested that about 20% of the students listened at levels that may have exceeded the 85 dBA action level. These high-volume level listeners included more males and more students of lower socioeconomic status compared to other groups. Although 2.3% of children had indications of hearing loss, hearing loss was not associated with personal music player usage (Twardella et al., 2017).

Personal Listening Devices in Background Noise

Williams (2005) examined the volume levels and listening habits of personal stereo devices of passers-by in real-life settings and the durations at which individuals are listening. Personal stereo player levels were measured on a KEMAR manikin with an artificial ear simulator at two locations: near the subway station and outside of a town hall. Males had greater average noise exposure levels than females, although average participant exposure level was 79.8 dB (below the risk level of 85 dB). Williams (2005) concluded that use of personal stereo devices alone did not increase the risk for a noise-induced hearing loss.

Ahmed et al. (2007) examined the relationship between the use of portable audio devices and hearing health of university students. Most students who owned MP3 players used the devices frequently (5-7 days per week) but listened at a mid-level volume. Approximately 13% believed they had a hearing loss which could be noise related. Objective data revealed that output was lowest when the background noise condition was the lowest and increased with the level of background noise. Ahmed et al. (2007) concluded that most students were listening at safe levels.

Kreisman (2014) examined the sound level and duration of MP3 player use by college students in five locations on a college campus (fitness center, library, quad, busy crosswalk, and student union). Results suggested significant differences in MP3 output level between most locations. Overall, 25% of participants exceeded the NIOSH permissible occupational noise levels, based on the measure of MP3 free-field equivalent (FFE) sound pressure levels in dBA FFE and reported estimated hours per day of MP3 use. However, no significant differences were found for percentage of students exceeding the NIOSH noise dose between the five locations.

Overhearing Music from Another Listener's MP3 Player

Advice about the sound levels emanating from a listener's headphones is relatively widespread. For example, MedlinePlus patient instructions on the webpage "Hearing Loss and Music" stated, "If you wear headphones, the volume is too loud if a person standing near you can hear the music through your headphones" (https://medlineplus.gov/ency/patientinstructions/000495.htm). This advice has been repeated despite the dearth of evidence that it is accurate.

Keith et al. (2008) measured output levels of nine MP3 players and twenty-nine earphones, including earbud earphones, supra-aural earphones, and circumaural earphones. Maximum free-field equivalent (FFE) output levels were 83.4 to 107.3 dBA, and depended on MP3 player, earphone sensitivity, tightness-of-fit in the ear, and recorded level of the music. In light of lay reports of subjective judgements of music listening that is too loud, Keith and colleagues measured the sound level of music observed by a "bystander" at 0.25 meters (10 inches) by presenting music at 94 dBA in the ear canal. The authors concluded that it is unlikely that a bystander would accurately judge an MP3 user's listening level, but the factors influencing such observations vary widely (most significantly, the level of background noise in the environment).

Weiner, Kreisman, and Fligor (2009) studied whether overhearing music from another person's MP3 player headphones indicated that the output level was loud enough to potentially damage the MP3 listener's hearing. Thirty participants with normal hearing were seated in a sound treated room surrounded by four speakers at 45, 135, 225, and 315 degrees azimuth. A probe microphone was placed in the participants' ear canal to measure the output level from the MP3 player. Participants were asked to select a random song on their MP3 player and set the volume to zero. The MP3 player screen was then covered and participants instructed to "adjust the volume to where you like it." A single observer (Weiner) measured the output levels in the ear canal and then moved 26-inches away, to determine whether the music could be overheard. These measures were repeated four times in thirty second intervals in four levels of randomized background noise: Quiet (ambient noise of 34 dBA) and 45, 60 and 75 dBA of pink noise. Results showed an average increase of 13 dB when listening in background noise and 26% of participants listened at levels above 85 dBA FFE in the highest background noise condition. A significant correlation was found between audibility and set listening levels in background noise suggesting that if music was overheard in noisy environmental settings, it was more likely that the music was set at a high intensity level. The sensitivity and specificity of "If I can hear it, that means it's too loud" was assessed and it was found that, in the quiet condition, sensitivity was 100%. However, as background noise increased, sensitivity decreased and the number of false negatives increased. In general, this study found that if you can overhear a person's MP3 player, it does not necessarily mean they are listening at an unsafe level (Weiner et al., 2009). The major limitation of this study was that it used only one observer, who was moving between MP3 listening levels and judging whether the music was audible. The authors stated that "A thorough study of this topic would use enough observers to provide adequate power and foil trials would be randomly interspersed with test trials" (Weiner et al., 2009).

Purpose

If being able to overhear another's music from his or her earphones is an adequate screening measure to determine if music listening is "too loud" (presents a risk to hearing health) or not, then the WHO-ITU Safe Listening Devices and Systems recommendation is unnecessary. We sought to apply a level of scientific rigor to the question of whether or not overhearing another's music from his or her earphones is adequate to screen for unsafe listening behavior. With these considerations in mind, the purpose of the present study was to examine the relationship between overhearing music from another person's MP3 player and the volume of that person's listening level, specifically whether the level exceeds the auditory risk criteria of 85 dBA FFE. We designed this study as a more thorough follow-up to Weiner and colleagues (2009).

METHODS AND MATERIALS

Participants

Fifty adults, 20 males and 30 females, ages 20-28 years (M= 23.5 years, SD = 1.7) participated in this IRB-approved study. All participants passed a hearing screening that is described in the Procedures section. Participants received a $10 iTunes gift card for their time.

Stimuli

Music

Five top selling songs from iTunes for the week of June 14, 2009 were selected as the music stimuli. The songs included "Boom Boom Pow" by the Black Eyed Peas, "Fire Burning" by Sean Kingston, "I Gotta Feeling" by the Black Eyed Peas, "I Know You Want Me" by Pitbull, and "Love Game" by Lady Gaga. A ten second clip was sampled from the chorus of each song. The song clips were equalized for overall RMS amplitudes using Adobe Audition 1.5 and were imported into iTunes. Table 1 displays the overall RMS values before and after equalizing RMS for both tracks of each song, as well as the relative peak amplitude of each song after equalizing RMS. The equalized song clips were then imported onto an iPod Touch with standard iPod earbuds.

Table 1.

Starting RMS amplitude, equalized RMS amplitude, and peak amplitudes after equalization of the five music stimuli song clips, in dB relative amplitude, as measured in Adobe Audition 1.3 software.

Song	Starting Left Track RMS	Starting Right Track RMS	Equalized Left Track RMS	Equalized Right Track RMS	Equalized Left Track Peak dB	Equalized Right Track Peak dB
IGF	-14.38	-14.29	-15.64	-15.55	-1.42	-1.32
BBP	-15.10	-15.00	-15.64	-15.55	-0.69	-0.59
FB	-13.31	-13.22	-15.64	-15.55	-2.52	-2.43
LG	-15.35	-15.25	-15.64	-15.55	-0.48	-0.38
IK	-15.64	-15.55	-15.64	-15.55	-0.23	-0.13

Note: IGF = I Gotta Feeling, BBP = Boom Boom Pow, FB = Fire Burning, LG = Love Game, IK = I Know You Want Me.

Recording what would be overheard

All measurements were carried out in a double-walled, sound-treated booth. A KEMAR mannequin was placed in a chair in the center of the test suite. A DPA 4006-TL omnidirectional microphone was placed at a distance of 28 inches at 0-degrees azimuth from KEMAR's head in order to simulate approximate distance of an individual passing by someone listening to the iPod. A distance of 28 inches was chosen, rather than the 10 inches used in Keith et al. (2008), as this distance is roughly the distance of two people standing or sitting in proximity without invading one another's personal space. The microphone was coupled to a DigiDesign Digi002 rack, which served as the external interface/router for the ProTools 7.3 software on the Macbook laptop. The EarPod earbuds of the iPod were placed in the ears of the KEMAR mannequin in order to record each song clip to simulate a listener using their iPod with earbuds. This one earphone was chosen for use in this study, as the iPod EarPod earphone is the highest selling earphone, capturing 60% of global market share (Counterpoint Research, 2019).

All samples were recorded in a quiet condition (ambient background noise of 31.6 dBA). The iPod volume (iPV) icon has 16 dashes which range from 0% to 100% iPV in 6.25% increments. For the study, song clips were recorded at every second dash on the iPV (0, 12.5, 25, 37.5, 50, 62.5,75, 87.5, and 100 percent). The peak output levels in dBA for each iPV of each song clip were measured via an IVIE IE-35 real-time analyzer/sound level meter, with a Type-I microphone located next to the recording microphone (see Figure 1). At 100% iPV, peak levels at the recording microphone ranged from 56.2 to 59.2 dBA. After the song clips were recorded, a 0.1-second 440 Hz tone was added at the beginning and ending of each song clip to signal when each clip starts and stops as some of the clips were inaudible (e.g. clips recorded at 0% iPV).

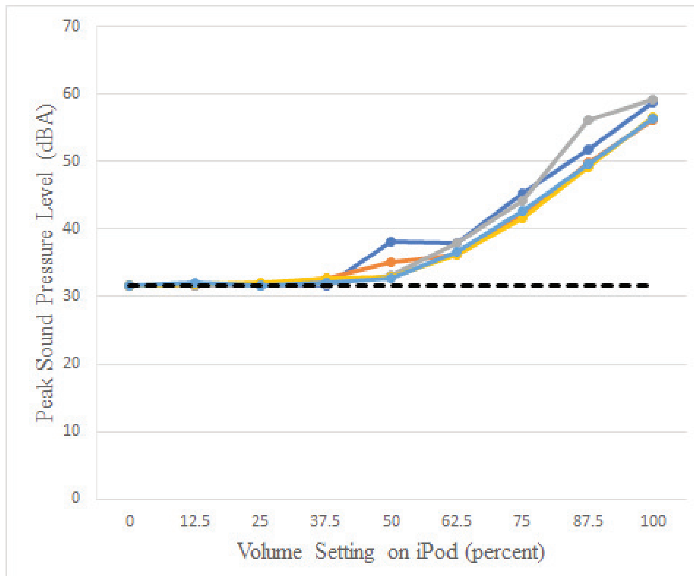


Figure 1. Peak Sound Pressure Level (in dBA) at the location of the observer of five songs at nine iPod volume settings (in percent). Note: Peak sound pressure level in dBA for five song clips playing through standard iPod earbuds on a KEMAR mannequin. Peak sound pressure levels were recorded in quiet at 2 feet four inches from KEMAR at 0 degrees azimuth to simulate the distance of a bystander. The dotted line represents the ambient background noise of 31.6 dBA.

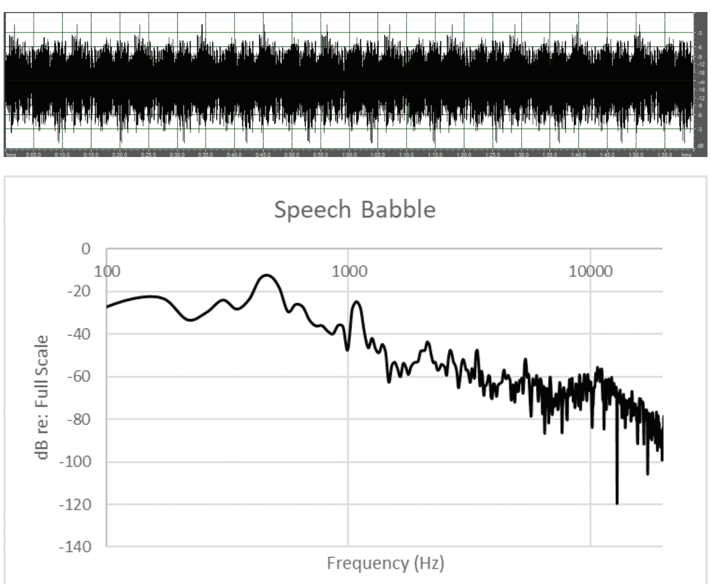
Background noise

The background conditions included quiet, three levels of pink noise (45 dBA, 60 dBA, and 75 dBA), and three levels of real-world background noise. These consisted of: 45 dBA speech babble copied from track 24 of the QuickSIN Speech-in-Noise version 1.3 CD (Etymotic Research Inc., 2006), concatenated to two minutes in length with no gaps in the signal; two minutes of 60 dBA restaurant noise from a recording at 44.1k Hz sampling rate inside a restaurant at lunchtime in downtown Boston, using a Shure SM58 microphone (Shure, Inc., Chicago IL) with XLR to USB adapter, recorded into Adobe Audition v1.5 software (Adobe Systems, San Jose, CA) in .wav file format; and two-minutes of 75 dBA airplane noise from a recording at 44.1k Hz sampling rate from inside of a flying Boeing 737 airplane using a Shure SM58 microphone with XLR to USB adapter, recorded into Adobe Audition v1.5 software in .wav file format. Two minutes of pink noise were generated using Adobe Audition v1.5 at -3 dB relative to full-scale.

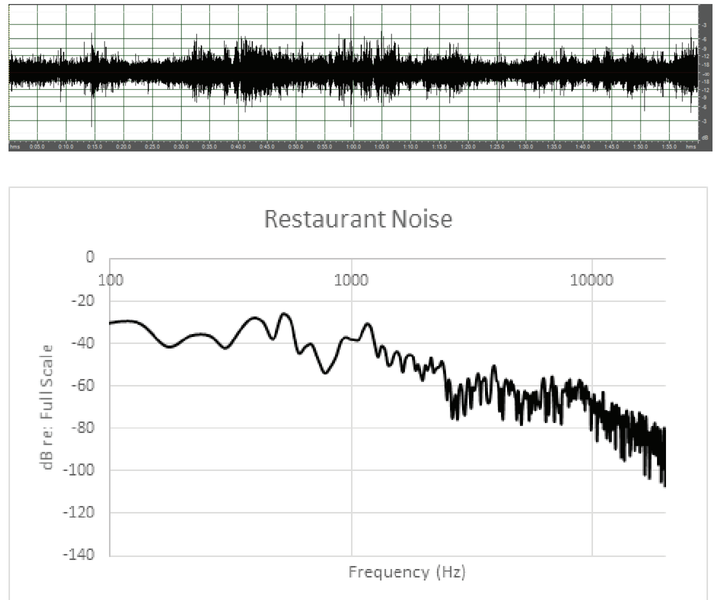
Pink noise was included as a background noise in the study as it is a non-kurtotic, gaussian noise, without peaks or gaps in the sound, as one would observe in real-world background noises; this ambient noise has also been used in several previous studies (Fligor & Ives, 2006; Weiner et al., 2009; Portnuff et al., 2013). Real-world background noise was included in this study to illustrate a few specific examples of a music-listener's environment. These real-world background noises should not be considered to generalize to all listening environments.

Speech-babble, restaurant noise, and airplane noise were equalized so that the highest peak in the signal was -3 dB relative to full-scale in Adobe Audition v1.5. Root-mean-square (RMS) level of speech-babble was 14 dB lower than the highest peak in the signal. Restaurant noise RMS level was 21 dB lower than the highest peak in the signal. Airplane noise RMS was 14 dB lower than the highest peak in the signal. Pink noise RMS level was 13 dB lower than the highest peak in the signal. Figures 2A, 2B, 2C, and 2D show the time-domain waveform and frequency response of each of the three real-world signals and pink noise.

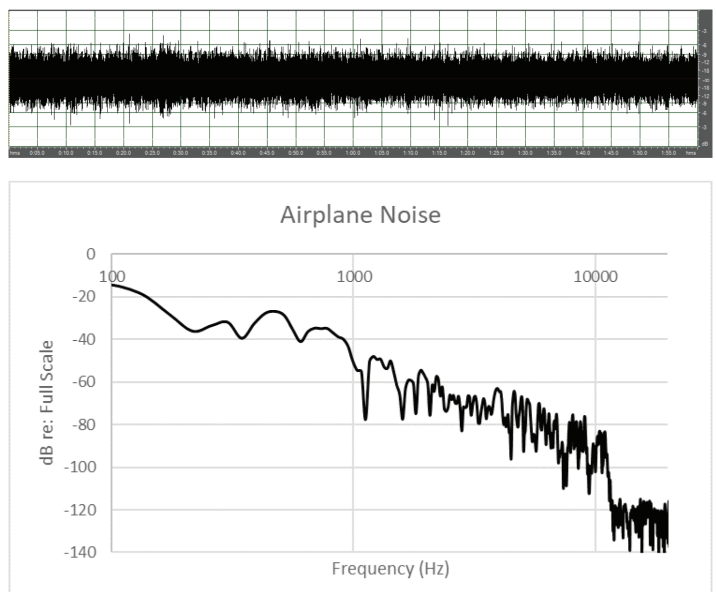
2A.



2B.



2C.



2D.

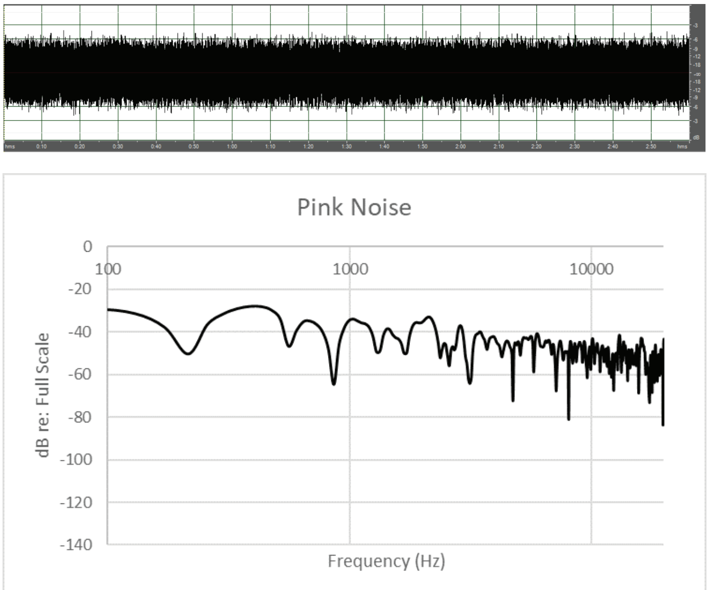


Figure 2. A. Time-domain waveform and frequency response of speech-babble signal. *B.* Time-domain waveform and frequency response of restaurant noise. *C.* Time-domain waveform and frequency response of airplane noise. *D*. Time-domain waveform and frequency response of pink noise.

Creating the final stimuli

Recordings were imported into separate tracks in ProTools 7.3 software, then sequenced to create the stimuli for each song, iPV, and background noise combination. The stimuli were routed from ProTools through a DigiDesign Digi002 rack to eight KRK Rokit Power 5 speakers. The speakers were arranged in an array 28 inches from the center of the participant's head at 0, 45, 90, 135, 180, 225, 270, 315 degrees azimuth. The song clip recordings were routed through the speaker at 0-degrees azimuth to simulate the sound coming from the iPod of a bystander, while the background noise was routed through the seven other speakers. Background noise levels were calibrated with an IVIE IE-35 real-time audio analyzer/Type-1 sound level meter placed in the center of the speaker array at the location of the listener's head.

Calculating FFE of the songs

Long-term average amplitude in dBA for each song clip at each iPV was measured in the ear canal of an investigator using an AudioScan Verifit system according to procedures described in ISO 11904-2 (2002), to determine sound levels at the iPod listener's eardrum. A probe tube was placed in the investigator's ear canal and a continuous sound level measurement was recorded for each song clip. Long-term average amplitude dBA of each clip was equated to FFE by subtracting out the eardrum transfer function in order compare the data to established damage-risk criteria (ISO 11904-1, 2002). Output at 100% iPV ranged from 91.6 to 95.8 dBA FFE. When presented at 100% iPV, all of the song clips had long-term average amplitudes greater than 85 dBA FFE, indicating potential risk for auditory damage. At 87.5% iPV, 4 out of the 5 song clips had long-term average amplitudes greater than or equal to 85 dBA. At 75% iPV and below, none of the song clips had average outputs greater than or equal to 85 dBA FFE. The means and standard deviations of the long-term average amplitude levels in dBA FFE are displayed in Figure 3.

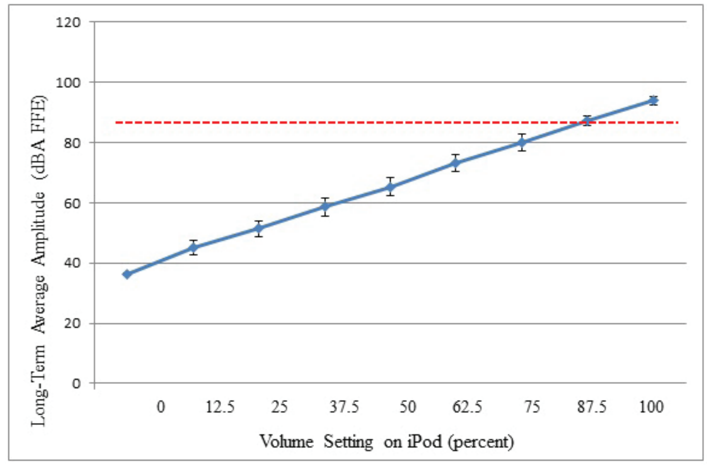


Figure 3. Means and standard deviations of long-term average amplitude levels in dBA FFE measures in the ear canal for five songs by volume setting. Blue diamonds indicate mean level for all songs at each volume setting. The dotted line at 85 dBA FFE indicates the hazardous listening levels.

Procedures

An audiologic screening was conducted on each of the participants to confirm normal hearing and normal middle ear function before participation in the study. An otoscopic examination was performed, followed by a bilateral tympanometry screening using a Madsen Otoflex 100 immittance bridge. A Type-A tympanogram (Jerger, 1970) was required. Pure tone air conduction testing was performed bilaterally using a pulsed 15 dB HL pure tone stimuli presented through a Grason-Stadler GSI 61 audiometer with E-A-R TONE 3A insert earphones. Thresholds were screened from 250 Hz through 8000 Hz including interoctaves. Following completion of the audiologic screening, the earphones were removed and the participant was seated in the center of the sound treated test suite. All participants passed the audiologic screening.

Written instructions were provided to the participant stating, "You will be listening to ten-second segments of noise and/or music. If at any time during the ten second segment, you hear the music, say 'yes.' The next sound clip will begin shortly thereafter. During some clips, you may not hear the music at all." After the participant indicated that he or she understood the instructions, the background noise levels were presented to the participant in order to familiarize the participant with the listening environment.

Due to the time required to complete testing, the testing was divided into two sessions lasting 60-90 minutes, with the first session being longer due to the hearing testing, and with a break in between the sessions. One session presented the clips in quiet and pink noise background noise, and the other session presented the clips in quiet and real-world background noise. The order of the test sessions was randomized for each participant. Within each session, the order of the clips was randomized for each participant, and each clip was presented once.

Results were analyzed via SPSS v. 17 using descriptive statistics. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each condition in order to determine whether audibility of the music over the background noise was indicative of hazardous listening levels. Receiver operating characteristic (ROC) curves were constructed for the two types of background noise (pink noise and real-world noise) to determine the optimal listening condition in which overhearing music from someone else's MP3 music is indicative that they are listening at hazardous levels.

RESULTS

Table 2 displays participant responses to each of the five song clips and each of the nine iPV settings for seven different background conditions. The number of participants who indicated the song clip was audible over the background noise are shown. We used a 50% criterion of the song clip being heard in a particular condition to be considered "audible." This 50% criterion was selected as the middle of a psychometric function. In the quiet background condition two of the five songs were audible at 25% iPV, and all songs were audible at 37.5% iPV and higher. In the 45 dBA pink noise background condition, one song was audible at 50% iPV and all songs were audible at and above 62.5% iPV. In contrast, in the 45dBA speech babble, one song was audible at 25% iPV, three songs at 37.5% iPV, and all songs at 50% iPV and higher. In the 60 dBA pink noise background condition, one song was audible at 75% iPV and all songs were audible at and above 87.5% iPV. In the 60 dBA restaurant noise condition, all songs were inaudible at and below 50% iPV, while all songs were audible at and above 62.5% iPV. In the 75 dBA pink noise background condition, none of the songs were audible regardless of iPV. In the 75 dBA airplane noise condition, one song was audible at 62.5% iPV and all songs were audible at and above 75% iPV.

Table 2.

Audibility Judgments of Song Clips Played on iPod at Nine Volume Settings in Quiet and Three Background Noise Levels with Pink Noise and Real-World Noise.

Noise Level	Noise Condition	Song	iPod Volume Level in Percent								
			0	12.5	25	37.5	50	62.5	75	87.5	100
Quiet	Quiet	1	1	6	22	92*	99*	97*	99*	100*	100*
		2	19	13	69*	95*	97*	97*	100*	100*	99*
		3	8	9	11	95*	99*	98*	99*	100*	99*
		4	6	14	51*	99*	96*	99*	100*	98*	97*
		5	10	13	43	97*	98*	97*	99*	100*	100*
45 dBA	Pink Noise	1	12	16	12	16	10	96*	94*	96*	96*
		2	10	8	8	18	46	94*	96*	100*	96*
		3	16	14	6	14	10	52*	96*	98*	100*
		4	10	10	8	6	86*	98*	100*	90*	98*
		5	6	6	12	6	20	96*	100*	98*	96*
	Speech Babble	1	4	2	0	32	96*	100*	100*	100*	100*
		2	0	28	88*	100*	100*	100*	100*	100*	100*
		3	6	6	2	32	98*	100*	100*	100*	100*
		4	0	0	4	72*	100*	100*	100*	100*	100*
		5	0	4	6	62*	98*	98*	100*	100*	100*
60 dBA	Pink Noise	1	18	10	12	10	20	12	16	80*	100*
		2	10	8	10	14	16	26	34	98*	98*
		3	16	18	6	8	18	20	6	54*	96*
		4	10	14	12	14	20	12	72*	94*	98*
		5	12	12	14	14	20	8	22	98*	96*
	Restaurant Noise	1	2	4	6	12	22	82*	100*	100*	100*
		2	6	6	10	14	44	98*	100*	100*	100*
		3	6	0	6	6	36	76*	100*	100*	100*
		4	4	4	8	8	42	90*	100*	100*	98*
		5	8	6	6	6	24	86*	100*	100*	100*
75 dBA	Pink Noise	1	4	6	10	10	0	8	4	10	8
		2	12	12	8	4	10	16	4	10	18
		3	16	4	12	6	6	6	14	4	4
		4	6	6	10	8	10	0	8	6	34
		5	6	8	2	16	10	2	12	8	2
	Airplane Noise	1	0	2	2	0	8	46	84*	100*	100*
		2	0	2	0	4	4	86*	100*	100*	100*
		3	0	2	6	0	4	22	98*	100*	100*
		4	2	0	2	2	2	42	100*	100*	100*
		5	6	2	6	0	0	48	100*	98*	100*

Note: The numbers are percent "yes" responses (n =100 for quiet condition; n = 50 for all other conditions). Asterisk (*) denotes song clip was audible during at least 50% of total trials in that condition.

In order to determine whether overhearing another person's iPod is a good predictor of whether he or she is listening at a hazardous level (a level equal to or above the 85 dBA FFE noise criteria), sensitivity, specificity, PPV, and NPV were calculated (see Table 3).

Table 3.

Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) for Each Noise Condition

Condition	Sensitivity	Specificity	PPV	NPV
Quiet	.99	.35	.28	.99
45 dBA Pink Noise	.97	.61	.38	.99
45 dBA Speech Babble	1.0	.47	.32	1.0
60 dBA Pink Noise	.95	.83	.58	.99
60 dBA Restaurant Noise	1.0	.63	.40	1.0
75 dBA Pink Noise	.11	.92	.26	.81
75 dBA Airplane Noise	1.0	.75	.50	1.0

Note: Sensitivity, specificity, PPV, and NPV based on 85 dBA cut-off value. Values for the quiet condition were calculated across both testing sessions.

Recall that if being able to overhear another's music from his or her earphones is an adequate screening measure to determine if music listening is "too loud" (presents a risk to hearing health) or not, then the WHO-ITU Safe Listening Devices and Systems recommendation is unnecessary. Sensitivity measures the proportion of actual positives that are correctly identified (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). For the purpose of this study, sensitivity is the percentage of people who indicated that they heard the song when the clip was actually playing at a long-term average amplitude level of 85 dBA FFE or louder. Sensitivity for songs played in ranged from .95 to 1.0 for all conditions except for the 75 dBA pink noise, in which sensitivity was only .11.

Specificity refers to the proportion of negatives that are correctly identified. In this study, this would be the percentage of people who did not indicate they heard the song when the clip was playing at a long-term average amplitude level below 85 dBA FFE. Specificity ranged from .35 in quiet to .92 in 75 dBA pink noise. In other words, songs playing at levels less than 85 dBA FFE were inaudible 92% of the time in the 75 dBA pink noise, while they were only inaudible 35% of the time in the quiet condition, indicating a high number of false positives (i.e., they heard the song even though it was below 85 dBA FFE).

Positive predictive value (PPV) uses both sensitivity and specificity and reflects the likelihood of a "disease" (here, a song clip playing with a long-term average amplitude level above 85 dBA FFE) when it is identified as being heard by the participants. Positive predictive value measures performance of a diagnostic method. For this study, PPV for the song clips being overheard by the observer show the probability that being able to overhear someone else's iPod is indicative of that person listening at hazardous level. Positive predictive value ranged from .26 in 75 dBA pink noise to .58 in the 60 dBA pink noise condition. In other words, even in the best condition, responses were only accurate 58% of the time. In general, PPV increased as background noise increased (except for the 75 dBA pink noise condition); however, PPV still remained poor. These results suggest that overhearing someone's iPod music is not a very accurate means of determining whether they are listening at a hazardous level.

Negative predictive value (NPV) shows the proportion of participants with a negative result who were correctly "diagnosed" (i.e., the proportion of participants who did not indicate that they heard the song when the clip was playing at a long-term average amplitude level below 85 dBA FFE) (Parikh et al., 2008). Negative predictive values ranged from .81 in 75 dBA pink noise to 1.0 in all three real world noise conditions. These scores indicated that when participants said they did not hear the clip in quiet, soft, or moderate pink noise or in the three real world background conditions, it was highly unlikely that the song clip had a long-term average amplitude level greater than or equal to 85 dBA. The NPV of .81 for the 75 dBA pink noise condition suggests that, compared to the other conditions, there is a greater likelihood that, although the song was not heard, it was being listened to at a hazardous level.

In order to determine the accuracy of a screening measure, a receiver operating characteristic (ROC curve) is constructed. The ROC curve provides a way to examine the accuracy of a diagnostic test and establish a threshold or cut-off for distinguishing between a positive or negative result (Rao, 2003). In our study, this test would be overhearing a song clip indicates a hazardous listening level.

Diagnostic tests typically involve a tradeoff between sensitivity and specificity (Rao, 2003). For example, if a threshold is set too low, results may show high sensitivity but low specificity, yielding many false positives. In our study, a false positive would be overhearing a song clip that has average amplitudes less than 85 dBA FFE criteria and therefore not a hazardous level. If a threshold is set too high, results may show high specificity but poor sensitivity. Rao (2003) reported that the best threshold or cutoff values have high sensitivity and low 1-specificity values. A low 1-specificity value indicates a low false positive (also known as false alarm) rate. When plotted on the ROC curve, this point will be located closest to the upper left corner of the graph. After all points are plotted, the area under the curve is examined. The larger the area, the more accurate the measurement tool is. A perfect measurement tool (100% sensitivity and specificity) would be located at the (0,1) intersection on the graph and have an area of 1.0 (Rao, 2003).

The ROC curves for pink noise and real-world noise are in displayed in Figure 4. The figure shows that test-performance was highest for the 60 dBA pink noise background condition, followed by the 75 dBA airplane noise for real world noise. In other words, overhearing music from someone's iPod in a 60 dBA pink noise background condition was the best predictor that the individual is listening at a level greater than or equal to 85 dBA. While the quiet condition has a high sensitivity value (99%), the specificity value was low (35%) indicating that when songs were played at levels less than 85 dBA, they were inaudible during fewer than half of the trials.

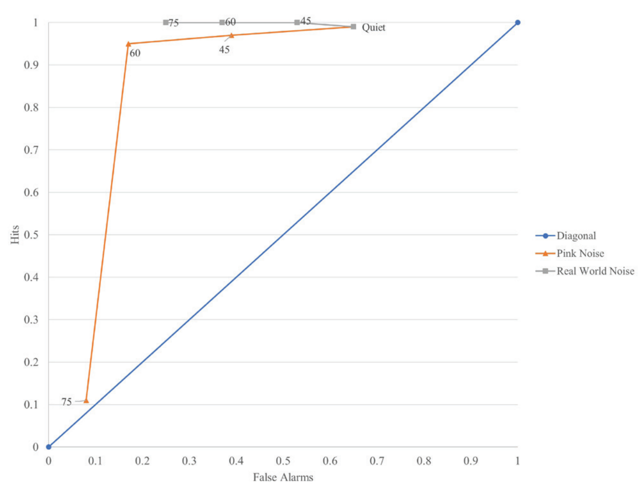


Figure 4. Receiver Operating Characteristic (ROC) curves for audibility of iPod playing at long-term average amplitude levels above 85 dBA FFE in different background noise conditions: Quiet (31.6 dBA of ambient noise), 45 dBA, 60 dBA, and 75 dBA of pink noise; and real-world noise of 45 dBA speech babble, 60 dBA restaurant noise, and 75 dBA airplane noise.

For pink noise, the thresholds or cutoffs increased as background noise levels increased from quiet to 45 dBA and 60 dBA of pink noise; however, when background noise increased to 75 dBA, the cutoff decreased significantly. Sensitivity was only .11 in this condition, while specificity was .92. These results indicated that the proportion of actual positives (song clips with long-term average amplitude levels greater than or equal to 85 dBA FFE) were only correctly identified (overheard by participants) 11% of the time. In contrast, for real world noise, sensitivity was near 100% for all conditions, while specificity decreased as background noise decreased.

Linear spectrum analyses of the 75 dBA pink noise and the five song clips were obtained in order to determine a probable cause for the lack of audibility of songs in that condition, in contrast to the airplane noise. It was determined that the song clips were completely masked by the noise. None of the song clips had peak sound pressure outputs above the energy in the 75 dBA pink noise and the signal to noise ratio was very poor. (see Figure 5).

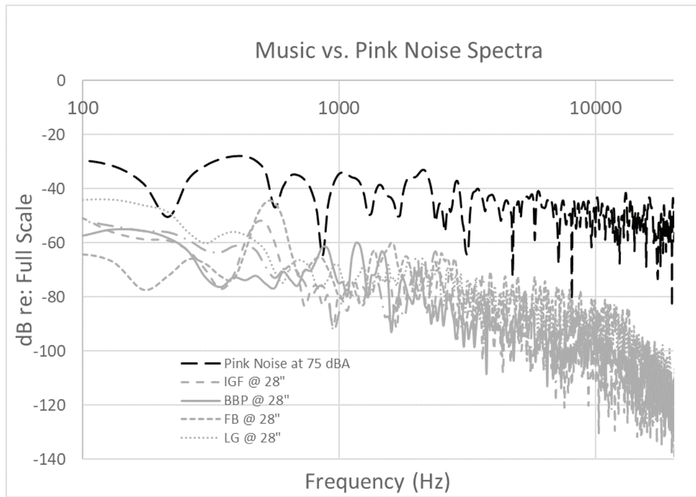


Figure 5. Frequency spectrum of pink noise at 75 dBA and frequency spectra of the five different songs based on sound level measured at 28 inches from the ear of the acoustical manikin. Note: IGF = I Gotta Feeling, BBP = Boom Boom Pow, FB = Fire Burning, LG = Love Game, IK = I Know You Want Me.

DISCUSSION

The purpose of this study was to rigorously apply test statistics to determine whether the ability to overhear another individual's iPod is indicative of a hazardous listening level for the iPod user that could potentially be harmful to hearing. The answer to the question "If I can overhear someone's iPod, does that mean they are listening at a hazardous listening level?" may not be a particularly clear cut one. Numerous factors influence audibility of an iPod including iPV, background noise, and song choice. Each of these factors must be taken into consideration.

Hazardous Listening Levels

When participants were listening for music clips in quiet, and in soft and moderate levels of pink background noise, clips playing at long-term average amplitude levels greater than or equal to 85 dBA FFE were audible more than half of the time to all participants. In a previous study by Weiner et al. (2009), when an observer was listening for music clips from an iPod in a quiet background condition and 45 dBA pink noise background condition, the majority of trials playing at or above the 85 dBA FFE level were also audible. Keith et al. (2008) predicted the same result, with 85 dBA FFE listening level being audible in 45 dBA of background noise. Weiner and colleagues found that in the 60 dBA pink background noise condition, a few trials where songs were playing at levels above 85 dBA FFE were inaudible to the observer. Findings from the current study also showed that during a few of the trials in the 60 dBA pink noise condition, not all participants overheard songs with long-term average amplitudes greater than or equal to 85 dBA FFE; however, the songs were audible to the majority of participants. Keith et al. (2008) predicted that bystanders on an idling city bus (60 dBA) would be able to hear higher frequency intermittent sound from earphones presenting music at 85 dBA FFE in the listener's ear.

Results from the current study revealed that when participants were listening for music in 75 dBA of pink background noise, none of the song clips were audible (in at least 50% of trials). Weiner et al. (2009) found that when an observer was listening for music stimuli in a 75 dBA pink noise background condition, some trials with long-term average amplitudes greater than 85 dBA FFE were audible to the observer. The audibility of certain clips was likely due to the fact that music stimuli levels were selected by participants and long-term average amplitudes for some songs exceeded levels of 100 dBA FFE. In comparison, the loudest song clips from the current study had maximum long-term average amplitude of 95.8 dBA FFE.

Volume Setting and Audibility

Peak sound pressure level (at the location of the observer) was measured for each song at iPV. Peak outputs varied from 31.6 dBA (background noise level) at 0% iPV to 59.2 dBA at 100% iPV. In the quiet background, no songs were audible at 0 and 12.5% iPV. Only two of the five songs were audible at 25% iPV, and all songs were audible at and above 37.5% iPV. These results suggest that songs are similarly but not equally audible across volume settings. Generally, audibility increased as iPV increased. Once audibility was achieved with a particular song in a particular condition, audibility was maintained with further increases in iPV.

These findings agree with previous findings from Weiner et al. (2009). Their results were that during all trials that were inaudible to the observer, music was playing at long-term average amplitude levels less than 85 dBA FFE. During trials where music was playing at levels greater than or equal to 85 dBA FFE, some song clips were audible to the observer while during other trials songs were not audible. Findings from the present study suggested that all songs at volume levels greater than or equal to 85 dBA FFE were audible to the majority of participants. Audibility of song clips played at volume levels less than 85 dBA FFE varied by the background noise condition. Findings from both studies demonstrated that audibility is not equal at all volume settings. These results are supported by the predictions of Keith et al. (2008), that audibility of the music to a bystander depended on the music being played (peaks relative to long-term average level), the level of the music, and the environmental noise level.

Background Noise

It is clear that the music was more audible in certain background conditions compared to the others. In the quiet background condition, participants overheard the song clips at low iPV (between 25% and 37.5%). These output levels were well below the 85 dBA FFE hazardous level. As background noise increased from 45 dBA to 60 dBA to 75 dBA, audibility of the songs decreased for both the pink noise and real-world noise. Audibility varied between pink noise and real-world noise at each background noise level. Compared to the pink noise, audibility for the real-world noise was achieved at a lower iPV. Recall that, in the loudest background condition of 75 dBA, no song was audible in the pink noise, while all songs were audible at and above 75% iPV in the airplane noise. This might suggest that environmental pink noise provided better test specificity (and better NPV) while the real-world background noises chosen in this study provided better test sensitivity (and better PPV).

When listening in a quiet, ambient environment, it is highly likely for one to overhear a nearby person's iPod, even if the person is not listening at a hazardous level. In the quiet condition, sensitivity was high; however, specificity was very poor due to the high number of false positives. In other words, it is common to overhear the music in a quiet setting even if the long-term average amplitude level of the song is below 85 dBA FFE, as indicated by the low PPV of .28 for this condition.

As background noise level increased, the PPV also increased, but not greatly. When music clips were presented 45 dBA pink noise background condition, PPV did increase in comparison to the quiet condition; however, PPV was only .38 for pink noise and .32 for speech babble. Positive predictive value was highest in the 60 dBA pink noise background condition. The PPV was .58, suggesting in this condition the test was accurate for overhearing songs with long-term average amplitude levels greater than or equal to 85 dBA FFE approximately 60% of the time. This finding suggests that if you overhear someone's iPod in a 60 dBA pink noise background condition, there is a higher probability that the person is listening at a hazardous level compared to overhearing the iPod in the two quieter background conditions. Comparatively, the PPV was lower at .40 in the 60 dBA restaurant noise condition. All song clips were inaudible in the 75 dBA pink noise background condition (using the 50% criterion), regardless of presentation level, and the PPV was only .26. Therefore, using this screening tool in the 75 dBA pink noise condition was practically useless, because even songs with long-term average amplitude levels greater than or equal to 85 dBA FFE were inaudible to the participants. In comparison, the PPV in the 75 dBA airplane noise was the .50, the second-highest in this study. Stated differently, in an airplane while 28 inches away, you could overhear someone's iPod with long-term average amplitude levels greater than or equal to 85 dBA half the time.

These findings are generally consistent with findings from Keith et al. (2008) and Weiner et al. (2009). That study determined that, as background noise increased, audibility of music to the observer decreased. Positive predictive value increased slightly and the number of false positives decreased as background noise increased. Weiner and colleagues also found that PPV was greatest in the 75 dBA pink noise background condition. In the current study, PPV generally increased as background level increased in both pink noise and real-world noise. The notable exception was when pink noise increased from 60 dBA to 75 dBA, the PPV decreased from .58 to .26. These results were likely due to the increased masking effects of the noise as pink noise exceeded even the highest peak sound pressure level by 15 dBA. The likely reason that our results differ from Weiner et al. is due to different methods; sound pressure of the music stimuli in Weiner et al. exceeded the song clips from our study by 5 dBA FFE or more. Data from our study suggests that audibility may increase as real-world background noise increases, which is similar to the trend that Weiner et al. found for pink noise. It should be stated that our findings are specific to the types of real-world noise in this study and caution should be used in generalizing the results.

Music Stimuli

It was determined that audibility varies slightly between songs. Even though songs were equalized for overall RMS amplitudes, there were still slight differences in their peak SPL. These differences caused some of the songs to be more audible than others in certain conditions. For example: 1) in the 45 dBA and 60 dBA pink noise conditions, song 4 was audible to participants at lower volume settings than the other songs; and 2) in the quiet condition, songs 2 and 4 were audible at 25 percent volume, while all other songs were inaudible at the same level.

Study Limitations

While this study was more rigorously designed than Weiner et al. (2009), a few limitations are worth noting. These include the choice of earphone, the distance of the listener, the hearing status of the overhearer, and the types of background noises evaluated. Caution should be used when generalizing our results. One limiting factor may have been that the study was carried out in a sound-treated test suite. Listener expectations within this synthetic listening environment may have contributed to a higher number of false positives in some of the conditions. The study only used ten-second samples of each song (a sample of the chorus) in order to save time and to reduce participant listening fatigue. There is a possibility that the selected ten-second clips did not capture the loudest peak SPL from the entire song, which could have potentially impacted results. Additionally, we only used one type of real-world background noise at each noise level. Results may vary with different types of real-world background noise. Finally, in order to control variability, we used only one type of earphone (the iPod EarPod). Therefore, results are limited to the EarPod and should not be generalized to other types of earphones or headphones, although we would expect most commercially-available earbud-style earphones to perform similarly.

Future Directions

While the current study used popular songs from iTunes as the music stimuli, it may be of interest to look at the audibility of songs from different genres of music; for instance, songs with higher peak-to-average sound levels would be more audible to a bystander, despite long term average levels being lower. In addition, the current study used an iPod Touch with standard iPod earbuds to assess audibility. In the future, it may be beneficial repeat this study with different styles of headphones, as some headphones attenuate background noise more than others (see, for example, Fligor & Ives, 2006; Keith et al., 2008). For example, if the earphone better seals the ear canal to attenuate background noise for the MP3 listener, less of the music sound may escape the ear canal. In such a case, a bystander may be prevented from overhearing the music even if it is being played at a hazardous level, thereby potentially changing the sensitivity and specificity.

Conclusion

The current study examined whether the ability to overhear someone else's iPod was indicative of that person listening at a hazardous level, or long-term average amplitude level greater than or equal to 85 dBA FFE. The results suggested that, in a quiet listening setting, it is highly likely that a nearby listener will overhear someone else's iPod; however, the chance that the person is listening at a hazardous level is only 28 percent, which is low. As background noise levels increase, the ability to overhear iPods generally decreases. Positive predictive value was higher in the louder conditions, indicating that if a person is able to overhear someone else's iPod in noisy listening setting, it is more likely that the person is listening at a long-term average amplitude level greater than or equal to 85 dBA FFE. In the 75 dBA pink noise condition, the noise completely masked all of the songs rendering them inaudible to the listeners regardless of the whether the levels exceeded 85 dBA FFE. It is necessary to keep in mind that the 85 dBA FFE noise exposure risk criterion is time-weighted. Therefore, individuals who are listening to music at long term average amplitude levels greater than or equal to 85 dBA FFE for long periods of time (8 hours or longer according to NIOSH standards) are at a greater risk for acquiring noise-induced hearing loss. Overall, these results suggest that whether or not a close bystander can overhear music from one's earphones is not a good screening measure for hearing loss risk. Technological interventions, such as those adopted by the WHO-ITU standard for safe listening devices and systems, are necessary to provide individuals with tools necessary to manage their risk for hearing loss from using MP3 players.

ACKNOWLEDGEMENT

Drs. Kreisman, Longenbach, and Blasi were formerly at Towson University, Department of Audiology, Speech-Language Pathology & Deaf Studies, Towson, MD. Dr. Fligor was formerly at Boston Children's Hospital/Harvard Medical School, Boston, MA.

REFERENCES

Ahmed, S., Fallah, S., Garrido, B., Gross, A., King, M., Morrish, T., …Pichora-Fuller, K. (2007). Use of portable audio devices by university students. *Journal of Canadian Acoustics, 35*(1), 35-54.

Bradley, R., Fortnum, H., & Coles, R. (1987). Research note: Patterns of exposure of schoolchildren to amplified music. *British Journal of Audiology, 21*, 119-125.

Catalano, P. J., & Levin, S. M. (1985). Noise-induced hearing loss and portable radios with headphones. *International Journal of Pediatric Otorhinolaryngology, 9*, 59-67.

Clark, W. W. (1990). Amplified music from stereo headsets and its effect on hearing. *Hearing Instruments, 41*(10), 29-30.

Counterpoint Research. (2019). Hearables Market Tracker for Q4 2018. Counterpoint Technology Market Research. Hong Kong.

Danhauer, J., Johnson, C. E., Byrd, A., DeGood, L., Meuel, C., Pecile, A., & Koch, L. L. (2009). Survey of college students on iPod use and hearing health. *Journal of the American Academy of Audiology, 20*, 5-27.

Etymotic Research Inc. (2006). QuickSIN Speech-in-Noise Test [CD]. Elk Grove, IL: Retvitronix.

Feder, K., Marro, L., Keith, S., & Michaud, D.S. (2013). Audiometric thresholds and portable digital audio player user listening habits. *International Journal of Audiology, 52*, 606-616.

Fligor, B. J., & Cox, L. C. (2004). Output levels of commercially available portable compact disc players and the potential risk to hearing. *Ear and Hearing, 25*, 513-527.

Fligor, B. J., & Ives, T. E. (2006). Does headphone type affect risk for recreational noise-induced hearing loss? Paper presented at the NIHL in Children Meeting, Cincinnati, OH. Retrieved from https://www.etymotic.com/media/publications/erl-0136-2006.pdf

Gopal, K. V., Mills, L. E, Phillips, B. S., & Nandy R. (2018). Risk assessment of recreational noise-induced hearing loss from exposure through a personal audio system-iPod Touch. *Journal of the American Academy of Audiology*. doi: 10.3766/jaaa.17140

Hellstrom, P., Axelsson, A., & Costa, O. (1998). Temporary threshold shift induced by music. *Scandinavian Audiology, 27* (Suppl 48), 87-94.

Hodgetts, W. E., Riegler, J. M., & Szarko, R. A. (2007). The effects of listening environment and earphone style on preferred listening levels of normal hearing adults using MP3 player. *Ear and Hearing, 28*(3), 290-297.

Hoover, A., & Krishnamurti, S. (2010). Survey of college students' MP3 listening: Habits, safety issues, attitudes, and education. *American Journal of Audiology, 19*, 73-83.

International Organisation for Standardisation 11904-1. (2002). Acoustics: Determination of sound emission from sound sources placed close to the ear: Part 1. technique using a microphone in a real ear (MIRE technique). Geneva, Switzerland: Author.

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolarygology, 92*(4), 311-324.

Keith, S. E., Michaud, D. S., & Chiu, V. (2008). Evaluating the maximum playback sound levels from portable digital audio players. *Journal of the Acoustical Society of America, 123*(6), 4227-4237.

Kohlrausch, A. (2007). The perceptual basis for audio compression. *Physics Today, 60*(6), 80-81.

Kreisman, B. (2014). Noise exposure estimates of college students using MP3 players in different locations. Poster presented at the American Academy of Audiology Conference, Orlando, FL.

National Institute for Occupational Safety and Health. (2018). Retrieved from http://www.cdc.gov/niosh/topics/noise

National Institute on Deafness and Other Communication Disorders. (2017). Listen up! Protect your hearing (infographic). Retrieved February 13, 2019 from https://www.nidcd.nih.gov/health/listen-infographic

Niskar, A. S., Kieszak, S. M., Holmes, A. E., Esteban, E., Rubin, C., & Brody, D. J. (2001). Estimaated prevalence of noise-induced hearing threshold shifts among children 6-19 years of age: The third national health and nutrition examination survey, 1988-1994, United States. *Pediatrics, 108*(1), 40-43.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology, 56*(1), 45-50.

Portnuff, C. D. F., Fligor, B. J., & Arehart, K. H. (2013). Self-report and long-term field measures of MP3 player use: How accurate is self-report? *International Journal of Audiology, 52*, S33-S40.

Rao, G. (2003). What is an ROC curve? *Journal of Family Practice, 52*(9), 695.

Pugsley, S., Stuart, A., Kalinowski, J., & Armson, J. (1993). Changes in hearing sensitivity following portable stereo system use. *American Journal of Audiology, 2*, 64-67.

Recommendation ITU-T H.870. Guidelines for safe listening devices/systems. (2018). Series H: Audiovisual and multimedia systems, E-health multimedia systems, services and applications – Safe Listening. International Telecommunications Union, Geneva, Switzerland.

Safe listening devices and systems: a WHO-ITU standard. (2019). Geneva: World Health Organization and International Telecommunication Union, 2019. Licence: CC BY-NC-SA 3.0 IGO.

Turunen-Rise, I., Flottorp, G., & Tvete, O. (1991). Personal cassette players ("Walkman"). Do they cause noise induced hearing loss? *Scandinavian Audiology, 20*, 239-244.

Twardella, D., Raab, U., Perez-Alvarez, C., Steffens, T., Bolte, G., & Fromme, H. (2017). Usage of personal music players in adolescents and its association with noise-induced hearing loss: A cross-sectional analysis of Ohrkan cohort study data. *International Journal of Audiology, 56*, 38-45.

Weiner, J., Kreisman, B., & Fligor, B. (2009). If I can hear their headphones, it's too loud, right? *Audiology Today, 21*(3), 45-51.

Williams, W. (2005). Noise exposure levels from personal stereo use. *International Journal of Audiology, 44*, 231-236.

Wong, T. W., Van Hasselt, C. A., Tang, L. S., & Yiu, P. C. (1990). The use of personal cassette players among youths and its effects on hearing. *Public Health, 104*, 327-330.