

An investigation of loudness effect on pitch and timbre discrimination

Wesley A. Bulla[†] and Song Hui Chon

Curb College, Belmont University, Nashville, TN, USA

[†]Corresponding author: wesley.bulla@belmont.edu

Published 16 December 2021; <https://doi.org/10.18061/FDMC.2021.0047>

Author video presentation and/or other conference material: <https://doi.org/10.17605/OSF.IO/MNHXG>

Abstract

This study was a first-step investigation of the efficacy of signal loudness when engaging in listening tasks. Based on physiological principles, pitch and timbre discrimination were hypothesized to be more accurate when stimuli were at low excitation levels. Discrimination tests were carried out at near audible threshold and at a robust somewhat loud listening level. Comparative metrics and non-parametric analyses indicated a consistent negative influence of signal strength on measured performance. While promising, more work is needed to determine the practicality and application of the findings.

KEYWORDS: *listening, cochlear tuning, discrimination, loudness, perception, pitch, timbre.*

Introduction

The ability to discriminate fine details in pitch and timbre are among the most important skills required of musicians, music producers, and audio engineers. While organizations such as the International Telecommunications Union (ITU) and the Audio Engineering Society (AES) have standardized calibration levels for equipment testing, currently, only *ad hoc* and often conflicting recommendations exist for proper listening loudness levels. Suggestions and documented common practices range from “conversation” level (Swedien, 2009) to as loud as 85 dBA SPL where the so-called “equal-loudness” effects (Fletcher & Munson, 1933; ISO, 2003; ITU, 2015) are believed to be beneficial in the discrimination process (Naqvi, 2013; Allen, 2014; Devine, 2017; McDonald, 2017).

Psychophysical evidence supporting claims of any case for the application of softer, intermediate, or louder listening levels when engaging in listening-oriented tasks is nonexistent. Additionally, the physiological mechanisms of cochlear tuning, loudness adaptation, and temporary threshold shift seem to be at odds with a tacit assumption that louder signals provide for easier, or more accurate, discrimination.

The Frequency Threshold Curve

Auditory coding begins when the inner hair cells in the Organ of Corti respond to fluid movement within the cochlea. Auditory nerve fibers are those afferent radial

neural filaments exiting the vicinity of the inner sensory cells of the cochlear organ (Rosenzweig, Leiman, & Breedlove, 1996). The Frequency Threshold Curve (FTC) of a cochlear nerve fiber is defined by its action-potential response (Hodgkin & Huxley, 1952) as charted against frequency sensitivity and signal strength (Matthews, 1999).

Investigations have shown that mammalian auditory nerve fibers exhibit a “V” shaped FTC where a fiber’s most sensitive frequency, designated its “characteristic frequency” (CF), resides near the threshold of detection (Pickles, 1984; Evans, 1993). Referring to Figure 1, note the response area of the FTC expands with signal strength where an auditory fiber normally associated with a given CF might generate an action potential in response to a stimulus frequency *not* normally associated with its designated CF.

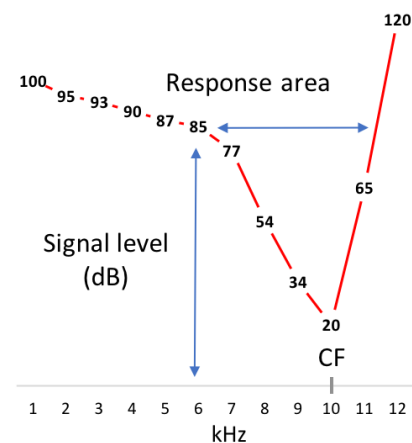


Figure 1. Stylized depiction of the FTC of a cochlear nerve fiber (adapted from Evans, 1993).

Of interest here is whether or not the variability in the FTC might interact with pitch, and by extension, timbre perception. If so, the perceptual outcome could manifest an increase in perceptual variability resulting in a decrease in discrimination accuracy commensurate with a rise in signal level.

Method

A four-part experiment was executed *via* headphone presentation in a semi-anechoic lab with a noise floor of



21 dBA (IEC, 2013). The complete experiment comprised pitch- and timbre-discrimination tests with soft and loud conditions. Headphone output was calibrated for two levels where “soft” was associated with listening near hearing threshold and “loud” was referenced to a robust signal of 83 dBA SPL. The loud listening level was chosen firstly because it was below the recognized action-level for mandated workplace hearing protections (OSHA, 2008); it matches well with popular press claims for so-called “best” listening levels when performing tasks such as blending musical instruments or making decisions regarding pitch and timbre; and finally, because it lies at a signal strength that is coincident with the observed broadening of the FTC.

The loud stimulus level was measured *via* ANSI/IEC 651 type-1 sound-level meter set to A-weighted, slow response (NMI, N.D.). Using the loud setting as a reference, the soft stimulus level was estimated *via* decrease of signal strength to just noticeable level within the noise floor of the room and headphone signal. The resulting signal was measured at -68 dB VU (15 dBA SPL). The just noticeable level was initially obtained by investigators’ audition and was confirmed *via* unanimous verbal acknowledgement from the subjects.

Stimuli

Stimuli were modeled on an aptitude test designed for fourth-grade assessment of basic music-oriented listening proficiency (Seashore, Lewis, & Sævetveit, 1960). Pitch test stimuli were seven 600 ms sinusoidal tones including 500, 508, 505, 504, 503, 502, and 501 Hz. The pitch stimuli set comprised sixty pairwise presentations where the 500 Hz stimulus was compared to each of the other six stimuli.

Stimuli for the timbre test were created by combining a 180 Hz sinusoidal tone with its first five harmonically related sinusoids (see Table 1). A resultant 1200 ms complex tone with phase-coincident harmonics at equal intensity was designated the “reference.” Timbral quality of the reference tone was varied by reciprocal adjustment of the intensity of the 3rd and 4th harmonic sinusoids.

Table 1. Harmonic tones in the timbre stimuli.

Variation	Sinusoidal (Hz)					
	180	360	540	720	900	1080
Reference	Equal intensity of all tones.					
v1	nc	nc	-10.0	+9.0	nc	nc
v2	nc	nc	-8.5	+4.0	nc	nc
v3	nc	nc	-7.0	+2.5	nc	nc
v4	nc	nc	-5.5	+1.2	nc	nc
v5	nc	nc	-4.0	+0.7	nc	nc

“nc” = no change of intensity

Five variations, each with greater differences in the reciprocal alternations were created. The timbre stimuli set comprised fifty pairwise presentations where the five variation stimuli (v1, v2, v3, v4, and v5) were each compared to the reference.

Subjects

Subjects were eleven graduate audio engineering students with ages clustered in the mid-twenties. All participants reported normal hearing and had completed a graduate course in ear-training for audio and music production. In a generalized sense, all subjects would be considered more experienced than “novice” but not yet “expert” listeners.

All subjects participated in all parts of the listening test. Subjects were randomly assigned to receive either the soft condition first with the loud condition following or vice versa. All four sessions were completed in one sitting. Instructions were given and sample stimuli were presented. However, there was no pre-test training. Each listening session lasted between six to seven minutes. Five-minute breaks were taken between sessions. The total testing session, inclusive of instructions, listening, breaks, and debriefing, lasted approximately 50 min.

Presentation

In each session, every possible pairwise evaluation was presented ten times in a double-blind randomized sequence. All pairs were presented with 350 ms between stimuli and four seconds between trials. In the pitch test, the first and second stimulus orders, (i.e., higher or lower) were balanced with five of each (per stimulus) randomized across trials. Subjects were asked to “indicate if the second tone in a pair was higher or lower than the first.” In the timbre test, the reference tone was presented first followed by one of the variations or a repeat of the reference. Subjects were then asked to indicate if the second tone in a pair “was the same or different from the first.” Subject responses were recorded manually *via* Scantron answer sheet (Scantron, N.D.) and collated electronically.

Results

This experiment employed simple discrimination tests where subjects’ per-trial responses were considered independent measures scored as either correct or incorrect. Repeated responses generated percent-correct measures for categorical and grouped analyses. Due to the binomial origin of data, scalar response metrics were transformed for non-parametric statistical

analyses (Peng, Lee, & Ingersoll, 2002). Correct response measures (%) are displayed for discussion.

Wilcoxon Matched-Pairs test indicated no significant difference between mean scores for soft VS loud condition order; therefore, data were combined without regard for condition order. *Shapiro-Wilk* tests rejected normality for all four data sets (pitch soft $W = 0.73, p < .001$; pitch loud $W = 0.86, p < .001$; timbre soft $W = 0.81, p < .001$; timbre loud $W = 0.93, p = .002$) and all data sets demonstrated a strong negative skew. Loud response data exhibited lower kurtosis and were more broadly distributed than soft condition data for both tests (figures 2 and 3, rightward panels).

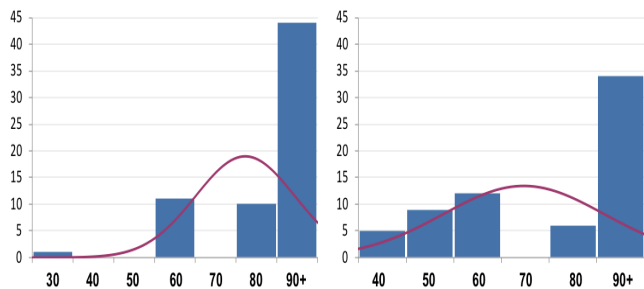


Figure 2. Percent correct score distributions for the pitch soft (left) and loud (right) data sets ($n = 66$).

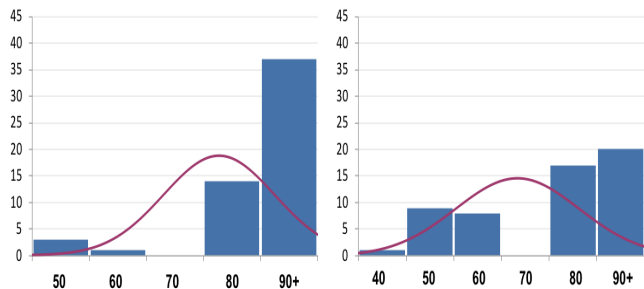


Figure 3. Percent correct score distributions for the timbre soft (left) and loud (right) data sets ($n = 55$).

Scores for both tests are plotted in figures 4 (pitch) and 5 (timbre). Here, each datum represents the mean of ten trials for each subject and each pair-wise stimulus combination. Comparative tests indicated significant differences ($Z = 4.58, p < .001, n = 66$) between means of the soft ($M = .90, SD = .08$) and loud ($M = .80, SD = .13$) pitch trials and between the soft ($M = .89, SD = .05$) and loud ($M = .73, SD = .14$) timbre trials ($Z = 4.26, p < .001, n = 55$) indicating loud condition group scores were significantly lower than soft condition group scores for both types of stimuli (see tables 2 and 3).

A non-parametric analysis of variance (ANOVA) indicated a significant effect of pitch difference in both

the soft ($X^2 (5, N = 11) = 34.5, p < .001, w = .62$) and loud ($X^2 (5, N = 11) = 31.6, p < .001, w = .57$) conditions suggesting pitch discrimination changed as differences (Hz) decreased. ANOVA found no influence of timbre variation (v1-v5) in the soft condition. Here, scores resided predominantly above 80% suggesting all variations were equally identifiable. In contrast, a significant effect was indicated for variations of timbre in the loud condition ($X^2 (4, N = 11) = 15.5, p = .003, w = .35$) where scores show increasing range and a greater number of incorrect judgements as differences between stimuli were decreased.

Table 2. Descriptive statistics of pitch discrimination performance.

Subject	Soft		Loud	
	M	SD	M	SD
S-01	0.95	0.08	0.87	0.20
S-02	0.95	0.08	0.87	0.15
S-03	0.93	0.12	0.92	0.13
S-04	0.92	0.12	0.90	0.20
S-05	0.92	0.13	0.87	0.22
S-06	0.78	0.12	0.52	0.13
S-07	0.92	0.13	0.90	0.15
S-08	0.72	0.23	0.73	0.16
S-09	0.92	0.10	0.85	0.16
S-10	0.93	0.12	0.82	0.10
S-11	0.93	0.12	0.62	0.17

Table 3. Descriptive statistics of timbre discrimination performance.

Subject	Soft		Loud	
	M	SD	M	SD
S-01	0.90	0.10	0.74	0.15
S-02	1.00	0.14	0.90	0.16
S-03	0.90	0.07	0.80	0.13
S-04	0.88	0.08	0.80	0.07
S-05	0.88	0.08	0.70	0.12
S-06	0.90	0.10	0.60	0.15
S-07	0.90	0.14	0.80	0.08
S-08	0.94	0.09	0.80	0.04
S-09	0.82	0.15	0.90	0.15
S-10	0.86	0.21	0.50	0.22
S-11	0.84	0.15	0.50	0.18

Grouped mean scores and standard deviations by stimulus are plotted in figures 6 (pitch) and 7 (timbre). Data here show increasing incorrect judgement trends consistent with those observed in the subject-by-stimulus scores shown in figures 4 and 5.

Discussion

Pitch test soft condition scores were above 90% for all but two subjects. In contrast, eight of eleven subjects

scored 80% or lower in the loud condition. Subjects' timbre scores exhibited similar trends. Notably, the magnitude of difference between the conditions was 60% larger for the timbre test as compared to the pitch test indicating potentially greater influence of signal strength interaction with timbre perception.

As expected, scores for both conditions show increasing incorrect judgments as differences between stimuli were reduced. Comparatively, loud condition incorrect judgments are noticeably more broadly spread over a greater range across more stimulus pairs (figures 4 and 5). Additional indicators lie in the consistent across-stimuli higher scores and lower variability for the soft condition as contrasted against the loud condition for both sets of stimuli (figures 6 and 7).

Score distributions, response metrics, and statistical analyses all suggest a consistent negative change in performance for the loud condition. Nonetheless, this study is far from complete. Firstly, treatment conditions employed represent two rather unrealistic boundaries that are fundamentally unworkable for common practice. The requirement for extremely low background noise makes the soft listening level difficult

to achieve and listening at or above the loud level is ill-advised as doing so will likely induce auditory fatigue and temporary threshold shift during prolonged exposures (Hirsh, Bilger, & Burns, 2016; Regers, 2016). Secondly, from a psychophysical perspective, while neural fibers and their corresponding CFs may be associated with tonotopic theories of pitch perception (Moore, 2003), the auditory system employs a variety of mechanisms (e.g., inhibition, dampening, amplification, and neural timing) in a system vastly more complex than a simple band-pass filter might suggest (Matthews, 1999; Howard & Angus, 2009). As a practical matter, one would expect cochlear loudness recalibration and adaptation (Scharf, 1997) to greatly influence the outcomes of any listening-focused task. An assumption of simple linearity of effect, i.e., "softer is better" would be, at least for now, vastly naïve.

This first experiment was a starting point in a broader investigation with many more intrinsic variables to consider and explore. While promising, clearly, more work is needed in order to fully understand the influence of loudness when making critical auditory-based decisions.

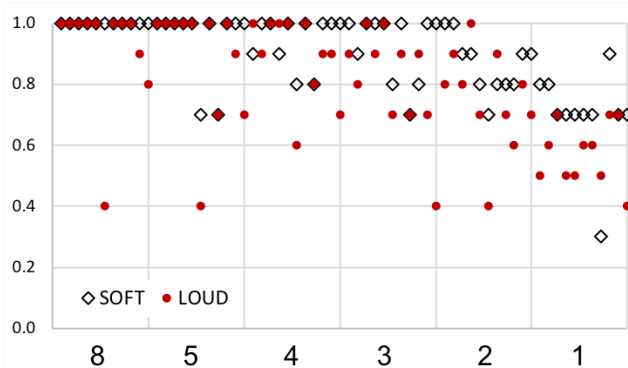


Figure 4. Subject scores by difference (Hz) for the soft and loud pitch tests.

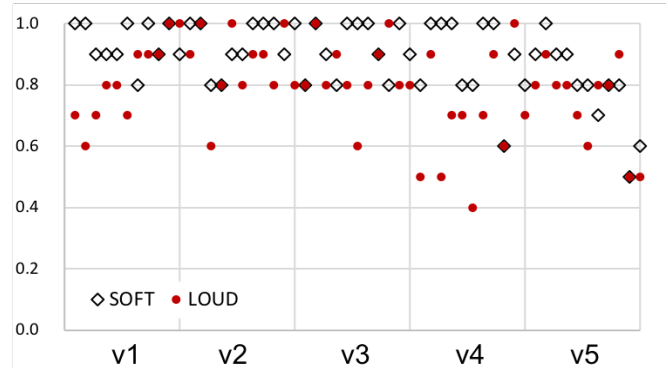


Figure 5. Subject scores by tonal variation for the soft and loud timbre tests.

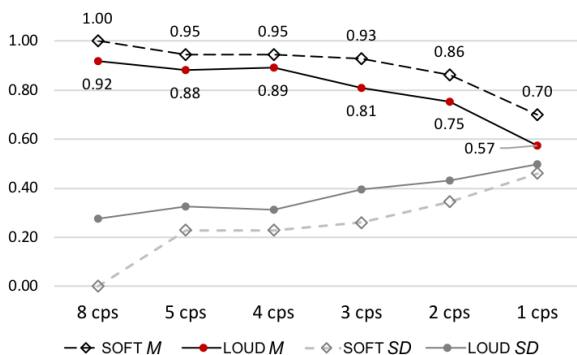


Figure 6. Group means (M) and standard deviations (SD) by stimulus for pitch trials.

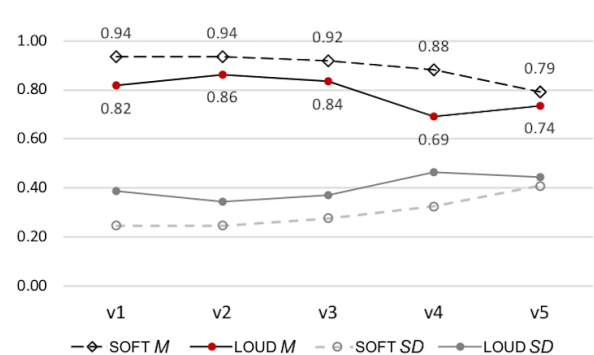


Figure 7. Group means (M) and standard deviations (SD) by stimulus for timbre trials.

Conclusion

While outcomes here indicate a consistent negative influence of signal strength on the accuracy of both pitch and timbre perception, more investigation is needed to determine the practicality and application of the findings. Specifically, next steps should examine changes of signal strength across the auditory bandwidth to determine whether or not observations here may reliably apply across a range of related stimuli and listening levels. Additionally, future work should be varied enough to determine if there is an optimum loudness range for engaging in a broad variety of listening tasks such as identifying differences of pitch, discriminating fine details in timbre, or the blending and balancing of musical instruments.

References

- Allen, R. (2014, November 9). *The Best Listening Level for Mixing*. (Rick Allen Creative Services, Inc. rickallencreative.com) Retrieved December 20, 2019, from <https://www.rickallencreative.com/production-buzz/best-listening-level-mixing>
- Devine, S. (2017, August 15). *Best Studio Monitoring Levels For Mixing*. (YouTube (at 5:23-5:41)) Retrieved December 20, 2019, from <https://www.youtube.com/watch?v=wJfcxPEjdN8>
- Evans, E. (1993). Basic physiology of the hearing mechanism. *AES 12th Conference: The Perception of Reproduced Sound. Paper #12-002* (Permalink: <http://www.aes.org/e-lib/browse.cfm?elib=6253>).
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am.*, 5(2), 82–108. <https://doi.org/10.1121/1.1915637>
- Hirsh, I. J., Bilgera, R. C., & Burns, N. W. (2016). Auditory-Threshold Recovery after Exposures to Pure Tones. *ASA 49th Meeting (M6)* (p. 1013). Acoustical Society of America.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4), 500-544. <https://doi.org/10.1113/jphysiol.1952.sp004764>
- Howard, D., & Angus, J. (2009). Chapter 2, Introduction to Hearing. In *Acoustics and Psychoacoustics (4th ed.)* (pp. 73-119). New York, NY: Focal Press. <https://doi.org/10.1016/B978-0-240-52175-6.00002-8>
- IEC. (2013). *IEC 61672-1 (2nd ed.) Electroacoustics - Sound level meters, ICS 17.140.50*. International Electrotechnical Commission.
- ISO. (2003). *ISO 226:2003 Acoustics - Normal equal-loudness-level contours (2nd ed.)*. Geneva, Switzerland: International Organization for Standardization.
- ITU. (2015, February 4). *Methods for the subjective assessment of small impairments in audio systems.* "Rec. ITU-R BS.1116-3 (02/2015) BS Series. Retrieved October 31, 2019, from International Telecommunications Union [<https://www.itu.int/rec/R-REC-BS.1116-3-201502-I/en>]
- Matthews, M. (1999). The Auditory Brain. In P. Cook (Ed.), *Music, Cognition, and Computerized Sound* (pp. pp. 1-10). MIT Press.
- Matthews, M. (1999). The Ear and How it Works. In P. Cook (Ed.), *Music, Cognition, and Computerized Sound* (pp. 11-10). MIT Press.
- McDonald, M. (2017, October 5). *How Loud Should You Mix?* (sweetwater.com) Retrieved December 20, 2019, from <https://www.sweetwater.com/insync/how-loud-should-you-mix>
- Moore, B. C. (2003). *An introduction to the psychology of hearing*. Academic Press (ISBN 0-12-505628-1).
- Naqvi, R. (2013, September 18). *PreSonus—How to calibrate your studio monitors*. (presonus.com PreSonus Audio Products) Retrieved December 20, 2019, from <https://www.youtube.com/watch?v=idGvZnSnPhs> (at 0:20-0:43)
- NMI. (N.D.). *Sound Level Meter Standards*. Retrieved from Noise Meters, Inc., Berkley, MI: <https://www.noisemeters.com/help/faq/standards/>
- OSHA. (2008). *Occupational noise exposure*. (URL: <https://www.osha.gov/enforcement/directives/04-00-004>): Occupational Safety and Health Admin., Federal Register, 29 CFR-2010, 1910-95.
- Peng, C. Y., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *J. Edu. Resrch.*, 96(1), 3-14. <https://doi.org/10.1080/00220670209598786>
- Pickles, J. O. (1984). Frequency threshold curves and simultaneous masking functions in single fibres of the guinea pig auditory nerve. *Hearing Research*, 14, 245-256. [https://doi.org/10.1016/0378-5955\(84\)90053-4](https://doi.org/10.1016/0378-5955(84)90053-4)
- Regers, S. N. (2016). Effects on Threshold Sensitivity of Exposing the Normal Ear to Pure Tones and White Noise over Sensation Levels of 10 db to 100 db. *ASA 49th Meeting (M7)* (p. 1013). Acoustical Society of America. <https://doi.org/10.1121/1.1918033>
- Rosenzweig, M. R., Leiman, A. L., & Breedlove, S. M. (1996). *Biological Psychology*. Sunderland, MA: Sinauer Associates.
- Scantron. (N.D.). *Insight 20/30 form no. 223127, GPAS-100Qx5*. store.scantron.com.
- Scharf, B. (1997). *Sequential Effects in Loudness*. Boston, MA (URL: <https://psychologie.lw.uni-leipzig.de/fechner/generalinfo/PDFs/BScharf.pdf>): Department of Psychology and Institute for Hearing, Speech, and Language, Northeastern University.
- Seashore, C., Lewis, D., & Saetveit, J. G. (1960). *Seashore Measures of Musical Talents (revised ed.)*. New York, NY: The Psychological Corp.
- Swedien, B. (2009). *Make Mine Music*. NY: Hal Leonard.