

Modeling the Individual Variability of Loudness Perception with a Multi-Category Psychometric Function

Andrea C. Trevino, Walt Jesteadt and Stephen T. Neely

Abstract Loudness is a suprathreshold percept that provides insight into the status of the entire auditory pathway. Individuals with matched thresholds can show individual variability in their loudness perception that is currently not well understood. As a means to analyze and model listener variability, we introduce the *multi-category psychometric function* (MCPF), a novel representation for categorical data that fully describes the probabilistic relationship between stimulus level and categorical-loudness perception. We present results based on categorical loudness scaling (CLS) data for adults with normal-hearing (NH) and hearing loss (HL). We show how the MCPF can be used to improve CLS estimates, by combining listener models with maximum-likelihood (ML) estimation. We also describe how the MCPF could be used in an entropy-based stimulus-selection technique. These techniques utilize the probabilistic nature of categorical perception, a novel usage of this dimension of loudness information, to improve the quality of loudness measurements.

Keywords Loudness · Categorical · Psychoacoustics · Normal hearing · Hearing loss · Suprathreshold · Perception · Maximum likelihood · Modeling · Probability

1 Introduction

Loudness is a manifestation of nonlinear suprathreshold perception that is altered when cochlear damage exists (Allen 2008). There are a number of techniques for measuring loudness perception (Florentine 2011); in this work we will focus on categorical loudness scaling (CLS). CLS is a task that has a well-studied relationship with hearing loss, uses category labels that are ecologically valid (e.g., “Loud”, “Soft”), can be administered in a clinic relatively quickly (<5 min/frequency), and requires little training on the part of the tester/listener. For these reasons, this task

A. C. Trevino (✉) · W. Jesteadt · S. T. Neely
Boys Town National Research Hospital, Omaha, USA
e-mail: Andrea.Trevino@boystown.org

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894, DOI 10.1007/978-3-319-25474-6_17

155

has been used in a number of loudness studies (Allen et al. 1990; Al-Salim et al. 2010; Brand and Hohmann 2002; Elberling, 1999; Heeren et al. 2013).

A listener's CLS function, generally computed from the median stimulus level for each response category, is the measure typically used in analysis. This standard approach treats the variability in the response data as something to be removed. We propose to instead model the probabilistic nature of listener responses; these probabilistic representations can be used to improve loudness measurements and further our understanding of the mechanisms underlying this perception. At each stimulus level, there is a probability distribution across categorical loudness responses; when plotted as a function of stimulus level, these distributions form a multi-category psychometric function (MCPF), modeling the statistics of the listeners' categorical perception. This model can be applied to listener simulations or when incorporating probabilistic analysis techniques such as estimation, information, or detection theory into analysis and measurements.

We describe one such application (i.e., ML estimation) to improve the accuracy of the CLS function. The ISO recommendations (Kinkel 2007) for adaptive CLS testing are designed to constrain the test time to a duration that is clinically acceptable. Due to the relatively low number of experimental trials, the natural variability of responses can create an inaccurate CLS function. Our proposed ML estimation approach is inspired by the work of Green (1993), who developed a technique for using estimation theory to determine the psychometric function of a "yes-no" task. We modify and extend this concept to estimate the MCPF that best describes an individual listener's categorical loudness perception.

In this paper, we (1) introduce the concept of the MCPF representation, (2) describe a parameterization method for the representation, (3) use principal component analysis (PCA) to create a representative catalog, (4) show MCPFs for NH and HL listeners, and (5) demonstrate how the catalog can be paired with a ML procedure to estimate an individual's MCPF.

2 Methods

2.1 *Multi-Category Psychometric Function*

A psychometric function describes the probability of a particular response as a function of an experimental variable. We introduce the concept of a multi-category psychometric function, which represents the probability distribution across multiple response categories as a function of an experimental variable.

In a MCPF, a family of curves demarcates the probabilities of multiple categories as a function of stimulus level. As an example, we demonstrate the construction of a hypothetical CLS MCPF for a NH listener, for an 11-category CLS scale, in Fig. 1. Figure 1a shows the listener's probability distribution across categories at 60 dB SPL. One can see that the loudness judgments are not constrained to one cat-

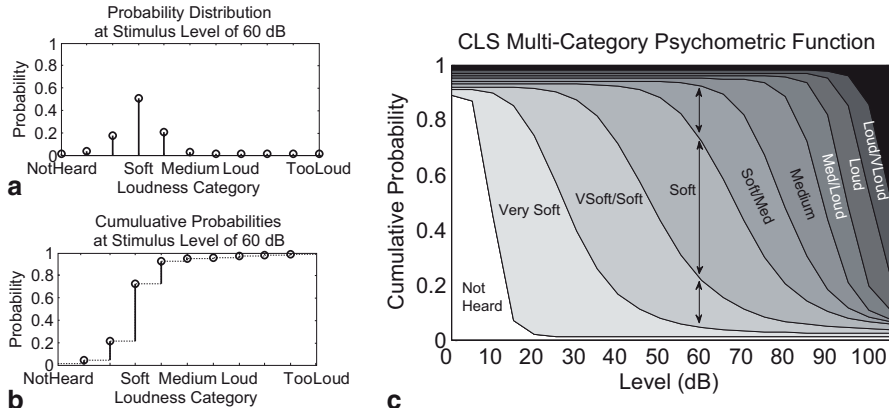


Fig. 1 Example MCPF, constructed from categorical probabilities. **a** Probability distribution across loudness categories at 60 dB SPL. **b** Cumulative probability function for loudness categories at 60 dB SPL; same data as **a**. **c** MCPF from 0 to 105 dB SPL. The vertical distance between curves represents the probability of each category. The 3 highest category probabilities at the 60 dB SPL stimulus level are marked by arrows. Maximum categories are not labeled

egory, but, instead, form a unimodal distribution. Figure 1b shows the cumulative probability density function for the data in (a). The MCPF (Fig. 1c) is constructed by plotting the cumulative distribution (marked by circles in (b)) as a function of stimulus level. In Fig. 1c, the probabilities of the top categorical response at 60 dB SPL are marked with arrows; the vertical distance between the curves matches the probabilities shown in Fig. 1a, 1b.

2.2 Parameterization

The MCPF curves are logistic functions that can be parameterized. A four-parameter logistic function $f(x | \theta)$, was selected to represent each of the curves of the MCPF. For our 11-category CLS scale, each MCPF, $F(x, i | \theta)$ consists of 10 category-boundary curves. Thus, the modeling implementation has 10 sets of parameters θ_i that define the 10 curves of each MCPF.

$$f(x | \theta) = D + \frac{A - D}{1 + (x | C)^B} \tag{1}$$

$$x = \text{stimulus level}, \theta = \{A, B, C, D\}$$

$A = \text{minimum asymptote}$, $B = \text{slope}$, $C = \text{inflection point}$, $D = \text{maximum asymptote}$

$$F(x, i | \theta) = f_i(x | \theta_i) = D_i + \frac{A_i - D_i}{1 + (x / C_i)^{B_i}} \quad i = 1, \dots, 10 \quad (2)$$

2.3 *A Representative Catalog*

We have developed a ‘‘catalog’’ of MCPFs to be representative of both NH and HL listener CLS outcomes based on listener data. Each function in the catalog has the same form (Eq. 2) and is defined by a set of parameters, $(\theta_i, i = 1, \dots, 10)$. A PCA analysis decomposes the primary sources of variability in the parameters. Let the matrix of all listener parameters be $X = \{\theta^1, \dots, \theta^M\}$, where M is the number of listeners, then the set of listener weightings, $W = \{w^1, \dots, w^M\}$, from the projections on the PCA eigenvectors v is $Xv = W$. Permutations of the sampled weightings were used to reconstruct the MCPFs that comprise the catalog. The superset of derived MCPF parameter sets that defines the catalog is denoted as Θ .

2.4 *Maximum-Likelihood Estimation*

The catalog can be used to compute a ML estimate of a listener’s MCPF. This may be particularly useful when the number of experimental observations is relatively low, but an accurate estimate of a listener’s loudness perception is needed, as is the case in most adaptive-level CLS methods that would be used in the clinic.

We denote a listener’s raw CLS data as (x_1, \dots, x_N) , where N is the total number of experimental observations. The likelihood, $\mathcal{L}(\cdot)$, of these observations is computed for each catalog MCPF, $F(x | \theta)$, maximizing over all potential parameter sets (i.e., all $\theta \in \Theta$). The maximization is computed over the functionally-equivalent log-likelihood.

$$\operatorname{argmax}_{\theta \in \Theta} \ln \mathcal{L}(\theta, x_1, \dots, x_N) = \operatorname{argmax}_{\theta \in \Theta} \sum_{i=1}^N \ln F(x_i | \theta) \quad (3)$$

Once the ML parameter set has been determined, it can be used to construct the listener’s MCPF, $F(x | \theta)$, via Eqs. 1 and 2.

Table 1 NH and HL listener characteristics. N is the number of participants. Age, audiometric threshold, and loudness discomfort level (*LDL*) are reported. Standard deviations are shown in parentheses. NA: tallies listeners that did not report a LDL from 0 to 105 dB SPL

	N	Age (yr)	Threshold, 1 kHz	Threshold, 4 kHz	LDL, 1 kHz (dB SPL)	LDL, 4 kHz (dB SPL)
NH	15	38 (± 10)	2 dB HL (± 5)	5 dB HL (± 5)	102 (± 4) [6 NA]	102 (± 4) [9 NA]
HL	22	53 (± 16)	31 dB HL (± 18)	51 dB HL (± 12)	102 (± 6) [3 NA]	100 (± 4) [11 NA]

2.5 Experiment

2.5.1 Participants

Sixteen NH listeners and 25 listeners with HL participated. One ear was tested per participant. NH participants had audiometric octave thresholds ≤ 10 dB Hearing Level; participants with sensorineural HL had octave thresholds of 15–70 dB Hearing Level (Table 1).

2.5.2 Stimuli

Pure-tone stimuli (1000-ms duration, 25-ms onset/off cosine-squared ramps) were sampled at 44100 Hz. CLS was measured at 1 and 4 kHz. Stimuli were generated and presented using MATLAB.

2.5.3 Fixed-Level Procedure

The CLS experimental methods generally followed the ISO standard (Kinkel 2007). Stimuli were presented monaurally over Sennheiser Professional HD 25-II headphones in a sound booth. Eleven loudness categories were displayed on a computer monitor as horizontal bars increasing in length from bottom to top. Every other bar between the “Not Heard” and “Too Loud” categories had a descriptive label (i.e., “Soft”, “Medium”, etc.). Participants clicked on the bar that corresponded to their loudness perception.

The CLS test was divided into two phases: (1) estimation of the dynamic range, and (2) the main experiment. In the main experiment, for each frequency, a fixed set of stimuli was composed of 20 repetitions at each level, with the presentation levels spanning the listener’s dynamic range in 5 dB steps. Stimulus order was pseudorandomized such that there were no consecutive identical stimuli and level differences between consecutive stimuli did not exceed 45 dB.

2.5.4 ISO Procedure for Testing ML Estimation

Five additional NH listeners, whose data were not used in the construction of the catalog, were recruited to complete an additional CLS task that used an adaptive stimulus-level selection technique, which conformed to the ISO standard. The adaptive technique calculated 10 levels that evenly spanned the dynamic range and presented these 3 times. For reference, for a listener with a 0–105 dB SPL dynamic range, the fixed-level procedure required 440 experimental trials while the ISO adaptive-level procedure required 30 trials.

3 Results

3.1 Individual Listener MCPFs

In each MCPF, the vertical distance between the upper and lower boundary curves for a category is the probability of that category. The steeper the slope of a curve, the more defined the distinction between categories, whereas shallower curves coincide with probability being more distributed across categories, over a range of levels. A wider probability distribution across categories indicates that the listener had more uncertainty (i.e. entropy) in their responses.

The NH and HL examples in Fig. 2 show some general patterns across both the 1 and 4 kHz stimuli. The CLS functions show the characteristic loudness growth that has been documented for NH and HL listeners in the literature. The most uncertainty (shallow slopes) is observed across the 5–25 CU categories. The most across-listener variability in category width is observed for 5 CU, which spans in width from a maximum of 50 dB to a minimum of 3 dB. A higher threshold and/or lower LDL results in a horizontally compressed MCPF; this compression narrows all intermediate categories.

The top two rows of Fig. 2 show data for two representative NH listeners, NH04 at 1 kHz (1st row) and NH01 at 4 kHz (2nd row). Listener NH04 had a low amount of uncertainty in their categorical response choices, with the most sharply-defined category boundaries at the lowest levels. Compared to NH04, listener NH01 has a wider range of levels that correspond to 5 CU (“Very Soft”) and a lower LDL, leaving a smaller dynamic range for the remaining categories. This results in a horizontally-compressed MCPF, mirroring the listener’s sharper growth of loudness in the CLS function. The bottom two rows of Fig. 2 show representative data for listeners with varying degrees of HL. Listener HL17 (3rd row) is an example of a listener with an elevated LDL. Listener HL11 (4th row) has LDL that is within the ranges of a NH listener; the higher threshold level causes the spacing between category boundary curves to be compressed horizontally. Despite this compression, this listener with HL has well-defined perceptual boundaries between categories (i.e., sharply-sloped boundary curves).

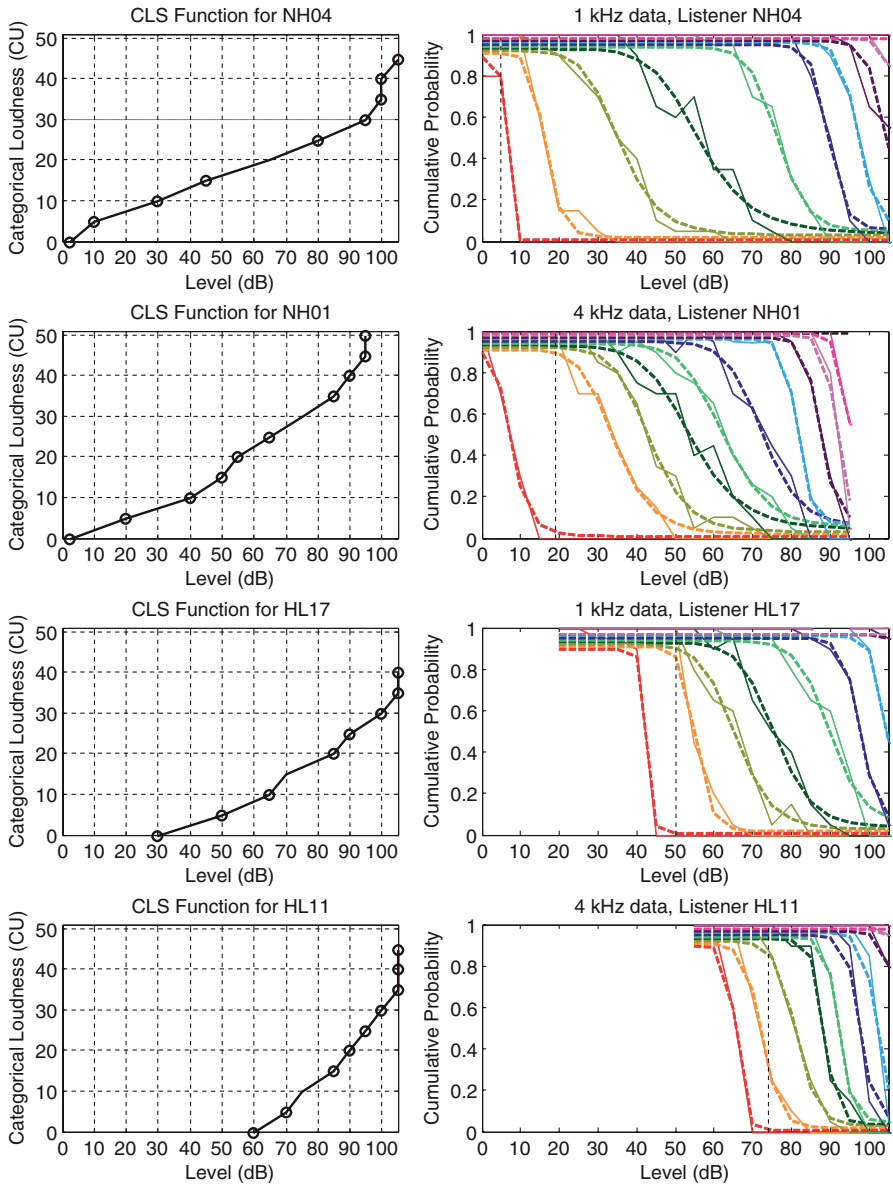


Fig. 2 The CLS function and corresponding MCPF for 4 listeners: (1st row) NH04, 1 kHz stimuli, (2nd row) NH01, 4 kHz stimuli, (3rd row) HL17, 1 kHz stimuli, (4th row) HL11, 4 kHz stimuli. The CLS function plots the median level for each CU. The MCPFs show the raw data as solid lines and the parameterized logistic fits as dashed lines. Each listener’s audiometric threshold is marked with a vertical *black* line

3.2 Construction of the MCPF Catalog

A PCA of the combined NH and HL data revealed that the first two eigenvectors were sufficient for capturing at least 90% of the variability in the listener parameters (θ). Vector weightings for creating the MCPF catalog were evenly sampled from the range (2 standard deviations) of the individual listener weightings. Permutations of 66 sampled weightings were combined to create the 1460 MCPFs that constitute the catalog.

3.3 Application to ML estimation

The MCPF catalog contains models of loudness perception for a wide array of listener types. Here, we demonstrate how a ML technique can be used with the catalog to estimate a novel listener's MCPF from a low number (≤ 30) of CLS experimental trials. As this adaptive technique has a maximum of three presentations at each stimulus level, a histogram-estimation of the MCPF from this sparse data can be inaccurate. The 50% intercepts of the ML MCPF category boundary curves may be used to estimate the CLS function (examples in Fig. 3). The average root mean squared error (RMSE) (fixed-level data used as baseline) for the median of the adaptive stimulus-level technique was 6.7 dB, and the average RMSE for the ML estimate, based on the same adaptive technique data, was 4.2 dB. The ML MCPF estimate improves the accuracy of the resulting CLS function and provides a probabilistic model of the listener's loudness perception.

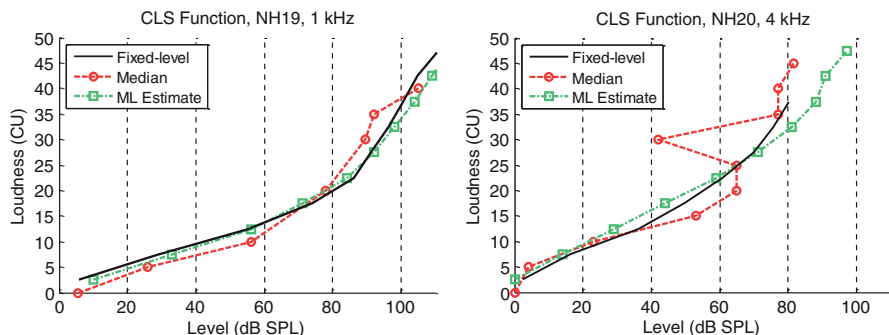


Fig. 3 Comparison of NH CLS results. The CLS function based on the fixed-level experiment is shown as a *black line*, representing the ‘best’ estimate of the listener’s loudness perception. The ISO-adaptive median is shown with *red circles*. The ML-based estimate is shown with *green squares*

4 Discussion

The CLS MCPF describes how categorical perception changes with stimulus level. The results demonstrate that loudness perception of NH and HL listeners is a random variable at each stimulus level, i.e., each level has a statistical distribution across loudness categories. The development of a probabilistic model for listener loudness perception has a variety of advantages. The most common usage for probabilistic listener models is to simulate listener behavior for the development of experiments or listening devices. Probabilistic models also allow one to apply concepts from detection, information, and estimation theory to the analysis of results and the methodology of the experiment.

In this paper, we show how a catalog of MCPFs can be used to find the ML estimate of a listener's MCPF, when a relatively small number of experimental trials are available. The ISO adaptive recommendation for CLS testing results in a relatively low number of experimental trials ($\approx 15\text{--}30$), in order to reduce the testing time and make the test more clinically realizable. Although this is an efficient approach, due to the low number of samples, the resulting CLS function can be inaccurate. The ML estimate of a listener's loudness perception based on this lower number of experimental trials is able to more accurately predict a listener's underlying CLS function, without removing outliers or using assumptions about the shape of the function to smooth the result. The MCPF catalog may be further exploited to develop optimal measurement methods; one such method would select experimental stimulus levels adaptively, such that each presentation maximizes the expected information. The ML MCPF estimate provides greater insight into the nature of the listener's categorical perception (Torgerson 1958), while still allowing for clinically-acceptable test times.

Acknowledgements This research was supported by grants from the NIH: T32 DC000013, R01 DC011806 (WJ), R01 DC008318 (STN), and P30 DC004662.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Allen JB (2008). Nonlinear cochlear signal processing and masking in speech perception. Springer handbook of speech processing. pp 27–60
- Allen JB, Hall JL, Jeng PS (1990) Loudness growth in 1/2-octave bands (LGOB)—a procedure for the assessment of loudness. *JASA* 88(2):745–753
- Al-Salim SC, Kopun JG, Neely ST, Jesteadt W, Stiegemann B, Gorga MP (2010) Reliability of categorical loudness scaling and its relation to threshold. *Ear Hearing* 31(4):567–578
- Brand T, Hohmann V (2002) An adaptive procedure for categorical loudness scaling. *JASA* 112(4):1597–1604
- Elberling C (1999) Loudness scaling revisited. *JAAA* 10(5):248–260
- Florentine M (2011). Loudness. Springer New York, New York
- Green DM (1993) A maximum-likelihood method for estimating thresholds in a yes–no task. *JASA* 93(4):2096–2105
- Heeren W, Hohmann V, Appell JE, Verhey JL (2013) Relation between loudness in categorical units and loudness in phons and sones. *JASA* 133(4):EL314–EL319
- Kinkel M (2007). The new ISO 16832 ‘Acoustics–loudness scaling by means of categories’. 8th EFAS Congress/10th Congress of the German Society of Audiology, Heidelberg
- Torgerson W (1958) Theory and methods of scaling. Wiley, New York