

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313 761-4700 800 521-0600

Order Number 9224825

**Loudness predictions using a physiologically based auditory
model**

Jeng, Patricia Sheau-Feng, Ph.D.

City University of New York, 1992

Copyright ©1992 by Jeng, Patricia Sheau-Feng. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

**LOUDNESS PREDICTIONS USING A
PHYSIOLOGICALLY-BASED AUDITORY MODEL**

BY

PATRICIA SHEAU-FENG JENG

**A dissertation submitted to the Graduate Faculty in
Speech and Hearing Sciences in partial fulfillment of
the requirements for the degree of Doctor of Philoso-
phy, The City University of New York.**

1992

© Copyright 1992 by PATRICIA SHEAU-FENG JENG
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

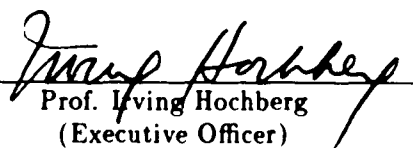
April 27, 1992

Date



Prof. Harry Levitt
(Chair of Examining Committee)

April 28, 1992



Prof. Irving Hochberg
(Executive Officer)

Dr. Joseph L. Hall

- Acoustics Research, AT&T Bell Laboratories.

Dr. Edward A. Cudahy

- New York League for the Hard of Hearing;

- Speech & Hearing, Graduate School of CUNY.

Dr. Bertrand Delgutte

- Dept. of Otolaryngology, Harvard Medical School;

- Research Laboratory of Electronics at M.I.T.;

- Eaton Peabody Lab., Mass. Eye and Ear Infirmary.

Supervisory Committee

The City University of New York

Abstract

LOUDNESS PREDICTIONS USING A PHYSIOLOGICALLY-BASED AUDITORY MODEL

by

PATRICIA SHEAU-FENG JENG

Adviser: Prof. Harry Levitt

Objectives of this dissertations were to predict loudness using a physiologically-based auditory model and to evaluate the model using loudness data. Loudness was studied because of its bearings in industrial and clinical applications and was used to evaluate the auditory model because of its dependence on harmonious functioning of many physiological mechanisms. The auditory model studied was the Ensemble Interval Histogram (EIH), which consisted of modules approximating functions of physiological units fundamental to audition. These units included middle ear, basilar membrane, inner hair cell, auditory nerves, and central auditory system. The loudness model (EIH-LM), based on the EIH model, was used to compute loudness values based on the assumption that loudness was proportional to the number of auditory neurons excited in the brain. Loudness data used for predictions included loudness functions (in sones), loudness level functions (in phons), and iso-loudness contours. For pure tones, published loudness data from Fletcher and Munson (1933), Hellman and Zwislocki (1976), and Scharf (1978) were used. For complex tonal stimuli, loudness data for four-tone complexes obtained by Scharf (1959) were used. In addition, an experiment was performed to obtain loudness level data for speech, narrow-band noise, and square wave stimuli. For each stimulus type, loudness levels were

measured for the test stimulus in quiet, partially masked by noise, and for the total loudness level of stimulus plus noise. Data was obtained for two types of reference stimuli and two methods of adjustment. Experimental results showed that (1) variations within a subject were smaller than those between subjects; (2) loudness levels obtained with different reference stimuli did not differ significantly; (3) a regression effect associated with two adjustment methods was observed; (4) loudness level functions for the square wave differed significantly from those for speech and narrow-band noise. Various model parameters were manipulated to examine the effect of these parameters on loudness predictions. The results showed that the accuracy of loudness predictions for both pure tones and complex stimuli depended on (1) the frequency range covered by cochlear filters; (2) low-frequency slopes of low-frequency filters; (3) the number of neurons activated within each cochlear channel; (4) neuron firing rate; and (5) integration time.

To
My Brother
Joseph Li-Ran Jeng

Acknowledgments

What I owe to each of these people who supported me through the process of acquiring a Ph.D. degree can never be repaid and the words come short of my sense of gratitude. I only wish to remember what was done for me and to repay by giving my best to the society that had given me this opportunity.

It was in 1985, Dr. Harry Levitt promised me that I would have no regret if I followed the route to research in Speech and Hearing instead of becoming a computer consultant. I took his advice, and here I am. There is no regret even though the road was rugged. The training was well worthy of the effort. It is not possible to thank Harry enough for giving me his undivided attention when I needed it. Without his assistance, there would be no conclusion to this work.

In 1987, Dr. David Berkley gave me the opportunity to work with Dr. Joseph Hall at Bell Laboratory at Murray Hill, a dream of many beginners in science. There, I was given the best support one could ask for in research. There, I was taken to a higher level of demand for quality in research. And there, I tasted the pleasure of research. I would like to thank Dr. Berkley for giving me this opportunity. I would like to thank Dr. Hall for supporting me in my work in many ways, and for being patient with me.

Bell Laboratory provided financial support for my work performed in this dissertation

and beyond.

I would like to thank all my subjects who gave me many hours out of their precious time and busy schedule to sit through tedious loudness matching sessions and gave me their useful comments.

I would like to thank Yariv Ephraim who gave me many useful professional and personal advice, suggestions, and encouragement.

I would like to thank Brent S. Townshend for his helps and critiques for my work, for many fun time at his and Sunil's house, and for those interesting tea time discussions in Cafe 1122.

Sunil Puria, my office mate, set an example before me in his hard working manner, in his generosity to help, and in his upbeat spirit. I really appreciate his fellowship through our Bell-Lab days and continuous encouragement afterward. I would also like to thank Nashie, a dear friend, for being there for me at my lowest point of the process and for seeing me through till the end.

My family has been to me as the source of energy compelling me to go forward. My father planted the seed for my sense of the world around me. My mother compelled me not to quit. My brother set a standard of excellence since my youth that I would follow him. My sisters told me I could do it. They surrounded me with love, faith, and prayers that gave me the courage to go on.

Finally, I would like to thank my dear husband with my deepest affection for going through heavens and hells with me, for staying out of my way, for cooking, washing dishes, and doing all sorts of chores, for being my friend, for being my teacher on demand, for being my critique, and for being my entertainment manager. He made it possible.

Contents

Abstract	v
Acknowledgments	vii
1 Introduction	1
2 Peripheral Auditory Physiology	8
2.1 Anatomy	8
2.2 Physiology of the Auditory System	9
2.2.1 Middle Ear	10
2.2.2 Cochlea	11
2.2.3 Primary Auditory Nerve	17
3 Loudness Calculation	21
3.1 History	22
3.1.1 Methods Developed by Fletcher, <i>et al.</i> at Bell Laboratories	22
3.1.2 Stevens' Method	39
3.1.3 Methods Developed by Zwicker <i>et al.</i>	47
3.2 Discussion	53
4 EIH Based Loudness Model	61
4.1 EIH Model	61
4.2 The Loudness Module in the EIH-LM Model	76
4.3 Twelve Versions of the EIH-LM Model	78
5 EIH-LM Prediction I	83
5.1 Predictions for Pure Tones	84
5.1.1 Loudness Functions	84
5.1.2 Loudness Level Functions	92
5.1.3 Iso-loudness Contours	96
5.1.4 Discussion	99
5.2 Predictions for Four-tone Complex	101
6 Loudness Matching Experiment	110
6.1 Introduction	110
6.2 Experimental Design	113

6.2.1	Procedure	113
6.2.2	Equipment	115
6.2.3	Stimuli	117
6.3	Results	118
6.4	Data Analysis	127
6.4.1	Test-retest Reliability	127
6.4.2	Analysis of Experimental Variables	132
6.5	Discussion	142
6.5.1	Sources of Subject Variations	142
6.5.2	Stimulus Level Adjustment	147
6.5.3	Averaged Loudness Levels	147
6.5.4	Regression Effect	152
6.5.5	Loudness Growth Curves	153
7	Model Predictions II	161
7.1	Predictions by EIH-Based Loudness Models	162
7.1.1	Predicted Detection Thresholds	162
7.1.2	Loudness Functions	166
7.1.3	Loudness Level Functions	176
7.2	Zwicker's Method of Loudness Calculation	184
7.2.1	Implementation of the Method	184
7.2.2	Loudness Function	188
7.2.3	Loudness Level Functions	190
7.2.4	Overall-loudness (O) Condition	194
7.2.5	Discussion	196
8	Discussion and Conclusion	197
8.1	Review of Loudness Prediction Methods	200
8.2	Loudness Matching Experiment	204
8.2.1	Loudness Measured in Quiet (Q)	207
8.2.2	Partial Loudness of Masked Stimuli (P Condition)	210
8.2.3	Overall Loudness of Masked Stimuli (O Condition)	211
8.3	Loudness Predictions by the EIH-LM Model	213
8.3.1	Summary of the Predictions by EIH-LM1	213
8.3.2	Effects of Various Parameters on Predictions	217
8.3.3	Complex Stimuli	222
8.4	Limitations of the EIH-LM Model	223
8.5	Recommendations for Further Research	227
8.5.1	Cochlear Filters	227
8.5.2	Level Detectors	229
8.5.3	Two-tone Suppression	229
8.5.4	Rate-level Function	230
8.6	Conclusion	230
A	Loudness Matching Experiment	232
A.1	Subjects	232

A.1.1	Subject Audiograms	233
A.2	Test Protocols	236
A.3	Loudness Matching Experiment Instructions	237
A.3.1	Instructions for Test Session 1: Session A of Quiet Condition (Q) . .	237
A.3.2	Instructions for Test Session 2: Session A of Partial-loudness Condi- tion (P)	237
A.3.3	Instructions for Test Session 3: Session A of Overall-loudness Con- dition (O)	238
A.3.4	Instructions for Test Session 4: Session B of Quiet Condition (Q) . .	238
A.3.5	Instructions for Test Session 5: Session B of Partial-loudness Condi- tion (P)	239
A.3.6	Instructions for test session 6: session B of Overall-loudness Condi- tion (O)	239
A.4	Thresholds of Six Test Stimuli	241
A.5	Loudness Matching Results	243
A.5.1	Mean Response Levels for Quiet Condition	243
A.5.2	Mean Response Levels for Partial-loudness Condition	244
A.5.3	Mean Response Levels for Overall-loudness Condition	245
A.6	Analysis of Variance for Experimental Variables	246
A.6.1	Significant Effects across All Subjects	247
A.6.2	Significant Effects Among Subjects	253

Bibliography

277

List of Tables

2.1	The values of parameters in the cochlear frequency-position function for human and for cat.	13
6.1	Stimulus Levels Used for Each Experimental Condition and Method of Adjustment.	112
6.2	Experimental Variables Used in the Loudness Matching Experiment.	128
6.3	Results of a Six-way, Repeated Measures Analysis of Variance of the Test-retest Standard Deviations.	129
6.4	Results of a Six-way, Repeated Measure Analysis of Variance on the Experimental Variables	133
6.5	Differences of the Averaged Detection Thresholds of the Six Stimuli from That of 1-kHz Pure Tone.	147
7.1	Predicted and Observed Detection Thresholds.	163
A.1	Identifying Information on Each Subject.	234
A.2	Test Protocols Using Latin Square Design.	236
A.3	Detection Thresholds of 5 Subjects for the Six Stimuli Used in the Loudness Matching Experiment.	241
A.4	Means and Standard Deviations of Response Levels for Quiet Condition Test Session A.	243
A.5	Means and Standard Deviations of Response Levels for Quiet Condition Test Session B.	243
A.6	Means and Standard Deviations of Response Levels for Partial-loudness Condition Test Session A.	244
• A.7	Means and Standard Deviations of Response Levels for Partial-loudness Condition Test Session B.	244
A.8	Means and Standard Deviations of Response Levels for Overall-loudness Condition Test Session A.	245
A.9	Means and Standard Deviations of Response Levels for Overall-loudness Condition Test Session B.	245

List of Figures

1.1	Schematic Diagram of the Physiological and Psychological Events Leading to Loudness Judgement, $Louder(S_1, S_2)$, for Two Input Stimuli, $S_1(t)$ and $S_2(t)$	2
2.1	Cochlear Maps for Human and Cat Based on Greenwood's Formula.	14
3.1	Loudness Level Data for 10 Pure Tones Measured by Fletcher and Munson, (1933).	27
3.2	Iso-loudness Contours Derived from Loudness Levels Data Measured by Fletcher and Munson.	28
3.3	Loudness Functions for 1-kHz Pure Tone and Their Slopes.	56
4.1	Block Diagram Depicting Functional Units of the Auditory System and Corresponding Elements of the EIH Model.	62
4.2	Original Cochlear Filters Used in EIH Model.	63
4.3	Sensitivity Thresholds of the Level Detectors in the EIH Model.	67
4.4	An Illustrative Diagram of EIH Model and Its Processing of a Pure Tone, $s(t)$	73
4.5	[.	75
4.6	New Set of Cochlear Filters Used in the EIH-LM Models.	79
5.1	Comparison of the Observed and Predicted Pure Tone Loudness Functions.	85
5.2	Slopes of the Pure Tone Loudness Functions as a Function of SPL are Shown.	88
5.3	Comparison of the Observed and Predicted Loudness Level Functions for Pure Tones.	93
5.4	Comparison of the Observed and Predicted Iso-loudness Contours for Pure Tones.	97
5.5	Re-plot of Scharf's Critical Band Data (1959, Fig. 3) Measured by Matching the Loudness of Four-tone Complex of Varying Bandwidth to a Comparison Tone at the Arithmetically Centered Frequency of the Complex.	102
5.6	Predicted and Observed Loudness Levels for Four-tone Complexes.	104
5.7	An Alternative Representation of Scharf's (1959) Data.	107
6.1	Block Diagram of the System Used in the Loudness Matching Experiment.	115
6.2	Measured Loudness Levels as a Function of Stimulus Levels in SL1k (re 1-kHz Tone Threshold) for the <i>Quiet</i> Condition (Q) Where the Reference Stimulus Was Adjusted in Level.	120

6.3	Measured Loudness Levels as a Function of Stimulus Levels in SL1k for the Quiet Condition Where the Test Stimulus Was Adjusted.	121
6.4	Measured Partial Loudness Levels as a Function of Stimulus Levels in SL1k for the Partial-loudness Condition (P) Where the Reference Stimulus Was Adjusted.	122
6.5	Measured Partial Loudness Levels as a Function of Stimulus Levels in SL1k Where the Test Stimulus Was Adjusted.	123
6.6	Measured Loudness Levels as a Function of Stimulus Levels in SL1k for the Overall-loudness Condition (O) Where The Reference Stimulus Was Adjusted.	124
6.7	Measured Overall-loudness Levels as a Function of Stimulus Levels in SL1k Where the Test Stimulus Was Adjusted.	125
6.8	Effect of Interactions AS and TAS for Each Subject on Test-retest Variability.	131
6.9	Effect of Stimulus Level, L, Experimental Condition, C, and Two-way Interactions, LC and AC.	136
6.10	Graphical Representation of TAC Interaction.	137
6.11	Interactions Involving the Variable S.	140
6.12	Original Loudness Growth Curves for Five Subjects.	144
6.13	Loudness Growth Curves Corrected for the Individual Stimulus Thresholds for Each of the Five Subjects.	145
6.14	Averaged Loudness Levels in Quiet; Stimulus Levels Specified in SL.	148
6.15	Partial-loudness Levels for Test Stimuli in Noise; Stimulus Levels Specified in SL.	149
6.16	Overall Loudness of Test Stimuli in Noise; Stimulus Levels Specified in SL.	150
6.17	Comparisons of Loudness Growth Curves for Speech (SP), Narrow-band Noise (NBN150), and Square Wave (SQW) in Three Test Conditions.	156
7.1	Loudness Functions Predicted by EIH-LM3, EIH-LM4, EIH-LM6, and EIH-LM7.	167
7.2	Loudness Functions Predicted by EIH-LM7, EIH-LM9, EIH-LM11, and EIH-LM12.	170
7.3	Predicted and Observed Loudness Level Functions in Quiet.	177
7.4	Comparison of Predicted Loudness Level Functions for EIH-LM4, EIH-LM6, and EIH-LM7 to the Observed Loudness Levels in the Loudness Matching Experiment.	180
7.5	Comparisons Between the Predicted and Observed Loudness Level Functions for SP+PN, NBN150+PN, and SQW+PN.	182
7.6	Block Diagram Depicting the Conversion Procedure Used to Convert Digitized Time Wave Forms of Sounds at Eardrum to Third-octave-band Spectra.	185
7.7	The Sound Pressure Level Transfer Function from Free-field to Eardrum at 0 Degree Azimuth. (Fig 9(a) in Shaw, 1974)	185
7.8	An Example of Specific-loudness Diagram of Speech and the Instantaneous Loudness of a Sentence as a Function of Time as Computed using Zwicker's Loudness Calculation Method.	187
7.9	Study of the Effect of FFT Window Length on the Loudness Calculated by Zwicker's Method.	188

7.10	Loudness Functions Calculated Using Zwicker's Method for Five Stimuli Used in the loudness matching experiment.	189
7.11	Comparisons of Loudness Level Functions Computed Using Zwicker's Methods with Those Observed.	191
7.12	Comparisons of the Observed and Predicted Loudness Level Functions for Three Test Signals Partially Masked by the Background Pink Noise.	193
7.13	Comparisons of predicted and Observed Loudness Levels Functions for the Overall Loudness of Three Test Stimuli Plus the Background Pink Noise.	195
8.1	Loudness Level Functions Measured Using PT and NBN120 as the Reference Stimulus for Each of the Test Stimulus in Each of the Experimental Conditions.	206
8.2	Loudness Level Function for SQW (the solid line fitted to the open circles and the asterisks) and for 100-Hz Pure Tone (dash-dotted line).	208
A.1	Pure Tone Detection Thresholds for Each Subject in dB SPL Between 100 and 8000 Hz.	235
A.2	Graphical Representation of LAC Interaction.	248
A.3	Graphical Representation of LTC Interaction.	249
A.4	Graphical Representation for LTA Interaction.	250
A.5	Interactions Among the Experimental Variables, L (Stimulus Level), T (Test Stimulus), A (Adjustment Method), and C (Experimental Condition).	252
A.6	Graphical Representation of CAS Interaction.	254
A.7	Graphical Representation of TAS Interaction.	255
A.8	Graphical Representation of RAS Interaction.	256
A.9	Graphical Representation of LAS Interaction.	258
A.10	Graphical Representation of TCS Interactions.	259
A.11	Graphical Representation of LCS Interactions.	261
A.12	Graphical Representation of LTS Interactions.	262
A.13	Graphical Representation of TACS Interaction.	265
A.14	Continuation of Previous Figure.	266
A.15	Graphical Representation of LACS Interaction.	267
A.16	Continuation of Previous Figure.	268
A.17	Graphical Representation of LTAS Interaction.	269
A.18	Continuation of Previous Figure.	270
A.19	Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S1 and S3.	273
A.20	Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S8 and S9.	274
A.21	Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S13 and S14.	275
A.22	Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S15 and S16.	276

Chapter 1

Introduction

Two main objectives of this dissertation are (1) to predict loudness using a physiologically based auditory model and (2) to evaluate the auditory model using psychophysical loudness data. Loudness, as a fundamental psychophysical attribute, has been studied extensively, because the understanding of how loudness perception arises is important for both industrial and clinical applications. The ability of the noise-control industry to predict loudness of various types of sound can provide a measure of the effects of noise on humans and thus provide improvement to environmental quality and prevention of noise-induced hearing impairment. The ability to predict loudness in the communication industry provides a way to quantify the effect of signal degradation. Clinically, a good understanding of loudness perception provides insights to loudness-related hearing impairment and thus makes possible the development of appropriate hearing aids and individually-tailored prescriptions to meet the needs of the hearing-impaired population. This dissertation provides an increased understanding of the limitations of this physiologically-based auditory model in predicting several aspects of loudness, including threshold, loudness growth, dynamic range, and relative loudness growth rates of various stimuli, and it quantifies the effects of

various parameters of the auditory model on its ability to predict loudness.

When a person makes a judgement that a sound, $S_1(t)$, is louder than another sound, $S_2(t)$, the physiological and the psychological events leading to the decision may be summarized by the block diagram in Fig. 1.1. The blocks represent (1) the pathway from the

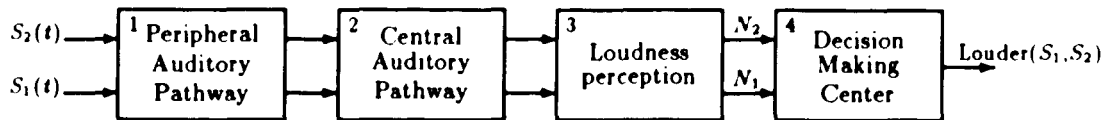


Figure 1.1: Schematic Diagram of the Physiological and Psychological Events Leading to Loudness Judgement, $Louder(S_1, S_2)$, for Two Input Stimuli, $S_1(t)$ and $S_2(t)$.

external ear to the first order auditory nerve; (2) the pathway from the cochlear nuclei to the auditory cortex; (3) the neural network in the brain that can recognize the *loudness* of the sound based on the neural information from the auditory cortex; and (4) the neural network in the brain that can make a decision based on the task instructions and the loudness perceived.

Psychophysics treats the functions in the four blocks as a single *black box* and investigates the response of the box based on variations of the input stimuli under the assumption that the response is affected by the input stimuli only. However, the response of any person does not depend solely on the physical stimulus. It also depends on the environment, and the internal physiological and psychological states of the person. Therefore, the person's response is often contaminated by other uncontrollable factors. Physiology, on the other hand, allows us more objectively to understand the *hardware* of the auditory system. At present, we have a more complete understanding of the peripheral auditory system than the rest of the pathway.

As early as 1921, driven by practical need, scientists at Bell Laboratories began to

investigate methods for predicting loudness¹ of complex sounds. The group, headed by Fletcher, started by using the intensity level (in dB) above threshold, i.e. sensation level, to represent loudness. Fletcher and Steinberg (1924) found that the total 'loudness' is the *sum* of the *cube root* of the energy in each frequency component across the speech spectrum². This was the first attempt at *loudness summation*. Soon, they found that sensation level in dB was inadequate, and the 'loudness' of each frequency band contributing to the total 'loudness' depended on both the threshold of the band and the total sensation level of the speech (Fletcher and Steinberg, 1924). The group went on to explore various phenomena in psychophysics relevant to loudness summation, such as loudness level functions (in phons), pure-tone iso-loudness contours, loudness function (in sones), critical bands, mutual masking, excitation patterns based upon the masking patterns, and related considerations. This body of data was then used to construct methods for loudness calculation (Fletcher and Munson, 1933, 1937; Fletcher, 1938, and 1950). Through the course of this development, the effort to understand the fundamentals of hearing became the driving force of their research rather than the need to estimate loudness for noise control. The principles of the physiology of the auditory system, known or hypothesized, were incorporated into their methods whenever possible. Unfortunately, this line of research that developed loudness summation methods has been largely ignored, but the principal of the results from Fletcher's group were adapted by other scientists.

Stevens (1956, 1961a, 1971) developed methods to sum loudness from 1-, 1/2-, or 1/3-octave band levels, based on the work by Fletcher's group, (Gates (1937), Beranek

¹The term *loudness calculation* has been used in international as well as U. S. standards. However, *loudness calculation* and *loudness prediction* are used interchangeably throughout this dissertation.

²In terms of pressure, the total 'loudness' would be the sum of the pressure to the 2/3 power. This is similar to the well-known Stevens' power law.

et al. (1951), Mintz and Tyzzer (1952), and Quietzsch (1955)). The parameters used in Stevens' method were obtained solely from psychophysical measurements. Stevens' method of loudness calculation was adopted as an American National Standard (ANSI S3.4/1980) and as an International Standard (ISO/R 532a - 1966(E)). Zwicker and his colleagues also developed a loudness summation model, indirectly based on some known physiological facts, but implemented the model using strictly psychophysical measurements (Zwicker and Feldtkeller, 1955; Zwicker and Scharf, 1965). The bases of their model were similar to those in the methods used by Fletcher *et al.*, except that the Bark scale, derived from critical bands, was used rather than the octave band as the loudness summation unit. The method by Zwicker *et al.* was adopted as an International Standard (ISO/R 532b - 1975(E)) for loudness calculation.

In the work presented in this thesis, a loudness model, constructed from a physiologically-based auditory model, is used to predict loudness for pure tones and complex sounds. The physiologically-based auditory model is called the *Ensemble Interval Histogram*, abbreviated as EIH. The loudness model, based on the EIH model, is called the *EIH-based loudness model*, abbreviated as EIH-LM. In addition to the loudness predictions made using the original EIH-LM model, various parameters of the EIH model are modified to give eleven variations on the EIH-LM model. These variations of the EIH-LM model are also used to predict a subset of the loudness data. The original and the eleven additional versions of the loudness model are abbreviated as EIH-LM1 to EIH-LM12. A comparison of the loudness predictions obtained from the model variations against each other and against human loudness data should improve our understanding of the effects of various model parameters on the loudness prediction. The prediction results are in turn used to evaluate

the EIH model and to gain insight into the strengths and the shortcomings of this type of model. Thus the second objective of the dissertation is fulfilled.

The EIH model is composed of a sequence of modules that approximate, to first order, functions of the physiological units that are fundamental to the peripheral auditory system (block 1 in Fig. 1.1). In addition, the EIH model simulates frequency extraction from the temporal synchrony present in the output of the peripheral auditory system. This frequency extraction belongs to the central auditory system. In the EIH-LM model, a loudness module is added to the basic EIH model in order to predict loudness, based on specific hypotheses of loudness coding (block 3).

Originated in the separate works by Allen (1985) and Ghitza (1985), and subsequent collaboration (Ghitza, 1986, 1987, 1988; Allen and Ghitza, 1990), the EIH model was used as a front-end processor for a speech recognizer. The EIH-based recognition results were very good, especially for the case of speech in the presence of noise (Ghitza, 1987, 1988). One of intents of this dissertation is to evaluate the EIH model for its strength and limitations.

There are many ways of evaluating the overall performance of a physiologically-based model. Two approaches found in the literature are: (1) use the model to predict physiological phenomena at either the acoustical-mechanical or neural level, and (2) use the model to predict psychophysical data. In this dissertation, the auditory model is used to predict loudness, a psychophysical phenomenon, and the prediction results are used to evaluate the model. The reasons for using loudness are as follows: Firstly, it is commonly believed that psychoacoustical loudness data largely reflects the capacity of the peripheral

transduction mechanism (Neff, 1961). Secondly, loudness perception depends on the harmonious functioning of all the physiological units of the peripheral auditory system, not on any one alone. Therefore, evaluation of the auditory model using loudness should provide a broader view of model's functions.

Recent work in physiology has greatly increased our understanding of the peripheral auditory system. In order to further the understanding of loudness perception, it is necessary to incorporate knowledge of auditory physiology into the loudness model. The EIH model by no means represents complete current knowledge of the auditory periphery. However, it contains some of the most basic elements of auditory physiology. Hence, the evaluation of the model using loudness prediction should provide some insights to Therefore, an objective of this dissertation is to study the limits of the EIH model for loudness prediction. The prediction results are intended to serve as a base line for further study of the roles of other auditory mechanisms in loudness perception.

The two objectives of this dissertation are really two sides of the same coin. Since both the auditory system and loudness perception are not fully understood, throughout the course of the study, the discussions are often presented from both approaches.

The organization of the dissertation is as follows: Chapter 2 provides a description of the peripheral auditory system. Chapter 3 reviews the major works in the history of loudness prediction since 1920. The review is by no means inclusive, but it presents the sources, to the best of the author's knowledge, of the basic ideas that are important to loudness prediction. Chapter 4 presents the EIH-based loudness model in four parts: The first part describes the EIH model and its underlying physiological bases. The second part presents the loudness module and the hypothesis of loudness coding used in the model to

predict loudness. The third part presents the parameter variations in twelve versions of the EIH-LM model. The fourth part defines terminology in the context of the loudness model. Chapter 5 presents loudness predictions for pure tones using EIH-LM1, EIH-LM2, and EIH-LM3, and for a four-tone complex using EIH-LM3. The predictions are then compared to observed loudness data already in the literature. Chapter 6 describes a psychophysical loudness-matching experiment in which loudness levels were measured for three complex stimuli at three sensation levels for three test conditions: loudness of signal in quiet, loudness of signal in noise, and overall loudness of signal plus noise. In Chapter 7, ten models, EIH-LM3 through EIH-LM12, are evaluated in terms of their ability to predict the loudness-level data of the experiment in Chapter 6. These loudness predictions are also compared to predictions based on Zwicker's methods for predicting loudness of signals in quiet, in noise, and overall loudness of the signal plus noise. Finally, Chapter 8 summarizes the work done in this dissertation and the major conclusions.

Chapter 2

Peripheral Auditory Physiology

In this chapter, peripheral auditory physiology is reviewed as a background for understanding the construction of the EIH model. Because the peripheral auditory systems of all mammals are largely similar, and because human physiological data are not readily available, the EIH model is constructed mainly from knowledge of nonhuman mammals, mostly of cat. The structure and the underlying principles of the model are discussed in Section 4.1.

2.1 Anatomy

Anatomically, the peripheral auditory system is commonly divided into three parts: the external ear, the middle ear, and the inner ear including the auditory nerve. The outer ear includes the pinna and the external ear canal. The middle ear is a cavity with a series of three small bones forming an ossicular chain which is attached to the ear drum (tympanic membrane) at one end, and to the oval window of the cochlea at the other end. The three bones are the malleus, the incus, and the stapes. The ossicular chain is also connected to the walls of the cavity via two small muscles, tensor tympani and stapedius muscle, and several ligaments. The former is connected to the malleus, and the latter is connected to

the stapes.

The parts of the inner ear that are relevant to auditory processes include the cochlea and the auditory nerve. The cochlea is similar to a fluid-filled tube coiled up into a snail-like form. The tube is longitudinally divided into three chambers: the scala vestibuli, the scala media, and the scala tympani. The three scalae are separated by Reissner's membrane and the basilar membrane. Sitting on the basilar membrane, along the full length of the scala media, is the organ of Corti, which houses the sensory cells called the inner hair cells (IHC) and the outer hair cells (OHC). The bases of the IHC are covered mainly with the synapses of afferent auditory nerve terminals that ascend to the cochlear nucleus. The OHC's are innervated mainly by efferent nerve fibers that project down from the crossed and uncrossed olivocochlear bundle (COCB and UCOCB) (Warr, 1978; Warr and Guinan, 1979).

2.2 Physiology of the Auditory System

The physiological functions of the peripheral auditory system are responsible for the transformation of the acoustic signal into neural firing patterns, through a sequence of complex processes. The processes include (1) band-pass filtering of the middle ear and level-dependent attenuation due to the acoustic reflex; (2) spectral analysis and compression along the basilar membrane; (3) nonlinear transduction of the mechanical motion of the basilar membrane into chemical and electrical signals in the hair cells; and (4) subsequent synaptic transduction of the electrical signals into the neural firing patterns along the nerve fibers. The mechanisms whereby frequency and amplitude information are extracted from sound and preserved in these transformation stages are crudely understood. Some

important details, such as the physics of level-dependent and other nonlinear operations are important current research topics. In order to understand the effects of these transformations on the frequency and the amplitude content of the acoustic signal encoded in various auditory components, and to be able to replicate these effects, it is important to model the cochlear and the middle ear transformations.

Throughout this work, binaural audition and external ear modeling are excluded, since most of the animal physiological data and most of the psychophysical data used for loudness predictions are monaural except in a few cases. Those cases will be specified when encountered. In addition, because most of the physiological and monaural psychophysical experiments calibrate the input acoustical stimuli with the sound pressure levels measured at the eardrum, only the middle ear and that part of the inner ear relevant to audition will be discussed.

2.2.1 Middle Ear

The tympanic membrane converts the rarefaction and the compression of the ear canal air pressure into mechanical motions which are then propagated through the ossicular chain to the oval window. There the piston motion of the footplate of the stapes transmits the wave pressure to the inner ear. The ossicular chain, in this sequence of mechanical motions, functions as a transformer that reduces the impedance mismatch between the air in the ear canal and the fluid behind the oval window in order to improve power transmission. By comparing the pressure measured in the ear canal to that behind the stapes, Nedzelnitsky (1980) obtained the pressure transfer function of the cat middle ear. The cat frequency response is like a band-pass filter with a maximum gain of about 30 dB at 1000 Hz and a drop-off of 8-10 dB per octave on the low-frequency side below 800 Hz and 3-6 dB per

octave on the high-frequency side above 2 kHz. The reduction of transmission at low frequencies has been ascribed to the elastic stiffness of the annular ligament, the tympanic membrane, the air wave inside the ME cavity (Guinan and Peake, 1967), and the decreasing input impedance of the cochlea below 1 kHz (Puria and Allen, 1991a). The reduction on the high-frequency side is due to the mass of the middle ear ossicles. The dip at 4 kHz, occurring only in cat, is due to the resonance in the bulla of the cat's ME cavity (Puria, 1991b). This dip does not occur in the sound transmission characteristic of the human ME, since the bulla cavity does not exist in human ME.

Sound pressure transmission through the cat's middle ear is linear up to at least 105 dB SPL across all frequencies (Guinan and Peake, 1967; Nedzelnitsky, 1980). However, this linear transmission can be modified nonlinearly by the acoustic reflex, which is induced either by very loud sounds or vibratory tactile stimulation of the ear. In laboratory animals, the acoustic reflex involves the contraction of both tensor tympani and stapedius muscles. In human, it is primarily the stapedius muscle that causes the acoustic reflex (Möller, 1983). The contraction of stapedius muscle pulls the stapes posteriorly and stiffens the annular ligament (Pang and Peake, 1986), which in turn reduces transmission below 1 kHz. The threshold of the acoustic reflex has been found to be about 80 to 85 dB SL when only the contralateral ear is stimulated by pure tones (methods commonly used experimentally and clinically). However, when both ears are stimulated, the thresholds can be as low as 65 to 70 dB SL (Möller, 1983, pp. 51).

2.2.2 Cochlea

The cochlea is the sensory organ where sounds are analyzed and transformed from mechanical wave motions into electrical neural signals. This process involves basically two

steps: basilar membrane transformation and hair-cell transduction.

Basilar Membrane (BM) The vibrations of the footplate of the stapes at the oval window displace the cochlear fluid and hence the BM, inducing a traveling wave of vibrations along the BM from base toward apex. This was first observed and reported by Békésy (1928, 1960). When the input is a pure tone, the travelling wave causes the BM displacement to reach a peak amplitude at a certain place on the membrane. The place where the BM displacement peaks is a function of frequency and, to a lesser extent, of amplitude. It is now widely accepted that the amount of displacement is a nonlinear function of the intensity (Rhode, 1971; Sellick *et al.*, 1982).

The important relation between stimulus frequency and the place of the maximum displacement on the BM, called the *cochlear map*, has been estimated using both psychophysical and physiological methods. Based upon Fletcher's suggestion (1953) that human critical bands may represent an equal distance along BM and that the cochlear map may be constructed by concatenating critical bands, Greenwood (1961a, 1961b, 1990) derived a formula relating frequency to place of maximum displacement on the BM for humans. Using the same formula with different parameters, he was also able to match the measured data for cats and many other species. Equation 2.1 is Greenwood's general form of the cochlear frequency-position function,

$$f(x) = A(10^{ax} - k) \quad (2.1)$$

where $f(x)$ is the frequency in Hz and x refers to place from the apex, either as a proportion of the total BM length, or as an actual distance in mm. The values for parameters A , a , and k depend on species; a can have either of two values depending on the units of x .

Table 2.1 summarizes the values of the parameters for human and cat. For human, a is

Species	A	k	a
Human	165.4	1.0	2.1 if proportion; 0.06 if mm
Cat	456.0	0.8	2.1 if proportion; 0.084 if mm

Table 2.1: The values of parameters in the cochlear frequency-position function for human and for cat.

2.1, if x refers to the place on BM, from the apex as a proportion of total BM length; or is 0.06 if x refers to the place from the apex in mm. For cat, a is 2.1 if x refers to proportion and is 0.084 if x is mm from the apex. The parameters for cat are the updated values given by Liberman (1982a) using direct physiological measurements on cats to relate the places on the BM to the characteristic frequencies (CF's) of tuning curves.

Fig. 2.1 shows cochlear maps for human and cat based upon Equation 2.1 with corresponding parameters specified in Table 2.1. The cochlear length for human is 35 mm and for cat 25 mm. The frequency range for human is from 20 Hz to 20 kHz and that for cat is from 100 Hz to 50 kHz. The top panel shows that human and cat have different cochlear maps. However, when the place on the BM is described as a percentage of the total length, the two cochlear maps are quite similar, with a geometrically fixed amount of frequency displacement, as shown in the bottom panel. This is consistent with our assumption that the mechanism of frequency analysis is similar in the two species.

As mentioned earlier, the displacement of the BM is a monotonic, nonlinear function of intensity. As the intensity increases, the total amount of the BM displacement increases both in height and in width of the peak. However, it is commonly believed that the amount of increase in the BM displacement decreases with the intensity, showing a compression effect at high sound pressure levels (Rhode, 1978; Sellick *et al.*, 1982). Due to measurement difficulties, accurate estimates of the displacement as a function of position (place) along

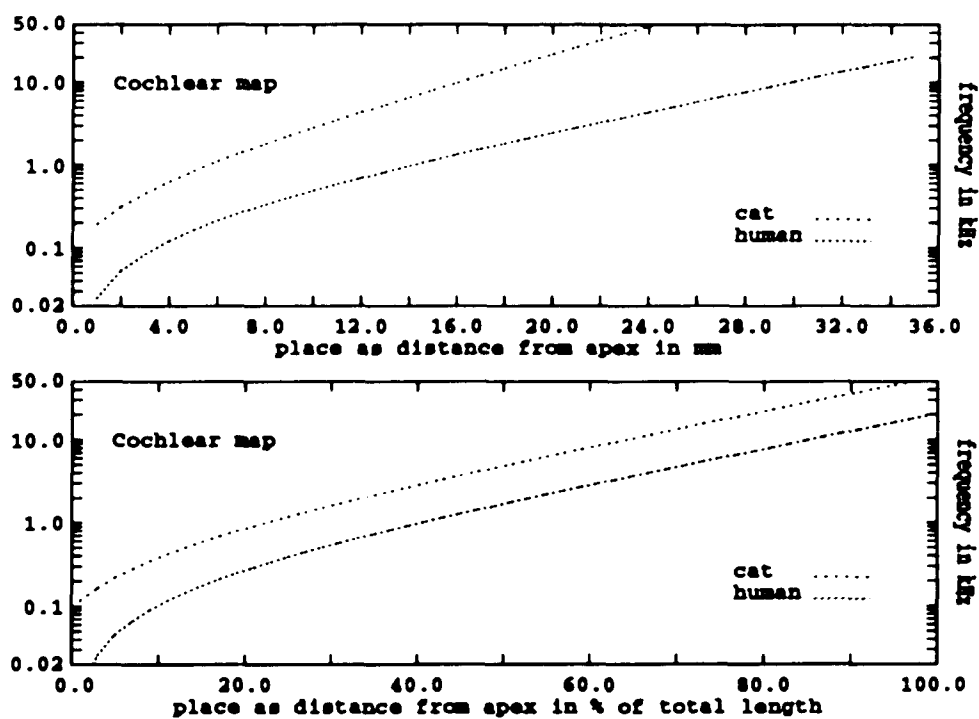


Figure 2.1: Cochlear Maps for Human and Cat Based on Greenwood's Formula. The place of the maximum displacement on basilar membrane as a function of frequency are shown for human and cat based on Greenwood's formula, Eq. 2.1. The ordinate is the frequency in kHz and the abscissa represents the place in mm, in the upper panel, and as the percentage of the total length from the apex in the lower panel.

the BM for pure tones are not available. Observations at one place, such as frequency responses or intensity function of the displacement, are available. Such data (e.g. Johnstone *et al.*, Fig. 4, 1986) show that the displacement is a linear function of the amplitude when the stimulus frequency is lower than the characteristic frequency (CF) or when the amplitude is at low intensity (Allen, 1989). If the stimulus frequency is greater than CF, or if the amplitude is at high intensity, the displacement becomes compressed. At CF, the displacement of the BM is approximately the square-root of the pressure at the eardrum. More data at different CF's would be necessary to accurately reconstruct the BM displacement for a pure tone.

When the input stimuli are complex, the travelling wave carries the displacement corresponding to each frequency component to its characteristic place on the BM, thereby decomposing the sound into its various frequency components. In addition, nonlinear interactions among the original frequency components of the sound generate other frequency components, called *distortion products*. These distortion products change the frequency composition of the basilar membrane response. The amplitude of the displacement at each place on the BM is further complicated by the interactions of the displacements of neighboring places on the BM. The total complexity of these interactions is not fully understood at this point. One form of interaction about which there is some knowledge is that of *two-tone suppression*. This interaction is discussed separately later.

Hair Cells The organ of Corti contains both inner hair cells (IHC) and outer hair cells (OHC). Physiological and anatomical evidence shows that sound transduction from mechanical motion into neural firings is through IHC's and their synapses at the IHC base. It is not clear how the OHC's function to modify this transduction process. However, it is

generally believed that OHC's play a role in the nonlinear compression of the BM response (Pickles, p. 75, 1988).

As the BM is displaced, the intracellular potentials (receptor potential) of the inner hair cells follow the deflections of the BM (Pickles, p. 59, 1988). These fluctuations of IHC receptor potentials have been characterized by their tuning curves, their intensity functions, and their input-output functions. The *tuning curve* of an IHC shows the sensitivity thresholds of the cell as a function of stimulus frequency. The *intensity function* for an IHC shows the receptor potential of the cell as a function of the stimulus sound pressure level (SPL). The *input-output function* shows the instantaneous value of the intracellular potential as a function of the instantaneous stimulus pressure during one pure tone cycle. Some researchers believe that IHC tuning curves and intensity functions have been shown to be similar to those measured on the BM (Rhode, 1978; Khanna and Leonard, 1982; Sellick *et al.*, 1982; Neely and Kim, 1983; Sellick *et al.*, 1983; Robles *et al.*, 1986). Others believe that there must be a *second filter* between BM and IHC responses, in order to provide neural tuning curves that are sharper than the mechanical tuning curves of the BM (Evans and Wilson, 1975; Hall, 1977a, 1981; Allen, 1977, 1980; Zwislocki, 1980; Zwislocki and Kletskey, 1980; Neely and Kim, 1986). This important disagreement greatly complicates the modeling problem.

An input-output function relates the instantaneous displacement of the IHC cilia to the instantaneous intracellular potential (sometimes called the alternating current or a.c. component) of the IHC response. When the BM moves upward (toward scala vestibuli), corresponding to negative-going pressure at the tympanic membrane, the cilia of the IHC are pushed radially outward due to the shearing motion between the tectorial membrane

and the BM. This motion reduces the electrical resistance of the cell membrane at the apex of the cell, increases the intracellular potential, and hence reduces the voltage difference inside and outside the IHC. This is called depolarization. The shearing motion induced by the positive-going pressure pushes the cilia of the IHC radially inward, causing the potential voltage to decrease, thereby hyperpolarizing the IHC. However, Cody and Russell (1987) show that the potential change in the depolarizing (excitatory) direction is much bigger and more gradual than in the hyperpolarizing (inhibitory) direction in IHC. On the contrary, the potential change is bigger in the hyperpolarizing direction than in the depolarizing direction in the OHC.

2.2.3 Primary Auditory Nerve

Innervating each IHC is a group of 20-25 primary auditory nerve fibers, each with a synapse at the base of the IHC. The caliber of these synapses for an IHC varies and is related to the sensitivity of the nerve fibers (Liberman, 1982b). The larger the synapse, the more sensitive the fiber and the lower the threshold. In each ear, there is a total of about 30,000 auditory fibers in the human auditory system and about 50,000 fibers in cat (Harrison and Howe, 1974).

Each fiber can be characterized by its firing rate in several ways. Firstly, the auditory fibers can be categorized into three groups based upon their spontaneous firing rates and their sensitivity threshold levels at the best (most sensitive) frequencies (Liberman, 1978). The *spontaneous rate* of a fiber is the rate of firing when the fiber is at rest, i.e. when not stimulated by an external stimulus. The *sensitivity threshold* level of a fiber is the sound pressure level of an input stimulus at which the firing rate increases from the spontaneous rate by a significant amount (the criterion rate). The three groups of fibers are classified

as high, medium, and low spontaneous rate fibers or as low, medium, or high threshold fibers. Liberman found that the sensitivity threshold of a fiber is highly correlated with its spontaneous rate. If a fiber has high spontaneous rate, its threshold is low, and vice versa if it has low spontaneous rate. The thresholds of low and medium threshold fibers tend to be more homogeneous, while those of high threshold fibers are not, but are spread over a 50-dB pressure range.

Secondly, the frequency selectivity of a fiber can be defined using the neural tuning curves at the sensitivity threshold, or by *iso-rate* and *iso-intensity* functions at suprathreshold levels. The *neural tuning curves* and the *iso-rate* functions express the sound pressure levels of various frequencies necessary to maintain the fiber at a fixed firing rate. The *iso-intensity* functions express the firing rate of the fiber for various frequencies at a fixed intensity level. Neural tuning curves are similar to IHC tuning curves and, in some cases (Robles *et al.*, 1986) to the BM tuning curves.

Thirdly, the firing rate of a fiber can be expressed as a *rate-level* function (or a *rate-intensity* function), i.e. the firing rate as a function of the intensity of an input stimulus. In general, the *rate-level* function of a neuron has a sigmoidal shape and has a limited dynamic range. The slope and the dynamic range of the function varies from fiber to fiber. The largest dynamic ranges are on the order of 40 - 50 dB. The smallest dynamic ranges are about 20 dB.

The relation between firing activity of a nerve fiber and the input stimulus over time can be observed in post-stimulus time (PST) histograms and period histograms. A *PST histogram* collates the number of spikes occurring within discrete time intervals according to their time of occurrence relative to the onset of the input stimulus. The PST histogram

for a tone burst typically shows a burst of firings at the onset of the stimulus. The rate of firing then declines to a stable level for the rest of the stimulus. The *period histogram* reveals the phase-locking property of nerve fiber to the period of the input stimuli. It does so by collating the number of spikes according to their occurrence time relative to a fixed point in a periodic cycle (e.g. the positive zero crossing of each cycle of the stimulus).

Temporal Synchrony Period histograms for low-frequency tonal inputs show that firings occur only on half of the cycle (excitatory direction), while spontaneous firings are suppressed during the other half of the cycle. Thus the firing patterns are phase locked to the input stimulus (i.e., they are temporally synchronized with the input stimulus). The phenomenon of synchrony may be another way of coding frequency information (Johnson, 1980), or of providing frequency selectivity beyond that of the mechanical frequency analysis of the BM. However, temporal synchrony occurs only for tones below 4 kHz. At and below the threshold level of a fiber, the firing pattern also shows the phase-locking property. This may be significant to the psychophysical acuity of hearing.

Two-tone Suppression The firing rate of a fiber evoked by a probe tone at the fiber's best frequency (probe) can be reduced by the presence of a second neighboring tone at an intensity level that is below the threshold of the fiber to that second tone (Sachs and Kiang, 1968; Liff and Goldstein, 1970; Arthur *et al.*, 1971; Abbas and Sachs, 1976; Hall, 1977b; Fahey and Allen, 1985). This is known as two-tone rate suppression. This effect has also been demonstrated at the IHC level by Sellick and Russell (1979) and on the basilar membrane by Patuzzi *et al.* (1984), and Robles *et al.* (1986a). A suppression curve for a fiber typically shows, as a function of frequency, the level at which the firing rate of the

fiber due to a probe tone is reduced by 20% or more. The suppression data shows that low-frequency tones are more effective in suppressing a probe of higher frequency than high-frequency tones in suppressing a probe of lower frequency. Sellick and Russel show that the response of an IHC to a probe at CF can be suppressed not only by a stimulus at a level that cannot excite that IHC, but also by a stimulus at a level that can excite that IHC when the stimulus is presented alone. Two-tone suppression gives us an appreciation of the complexity of sound transduction in the auditory system for complex stimuli.

Chapter 3

Loudness Calculation

A great amount of work has been done to understand how various auditory parameters contribute to the perception of the loudness of a sound and to derive a simple loudness prediction method for practical application. Efforts to find a suitable measuring unit to quantify and a reliable method to predict loudness of complex sounds can be traced back to the early 1920's. Because of the need to control noise and to quantify speech signal loss through various communication systems, scientists at Bell Laboratories initially (in 1921) used intensity in decibels (dB) above the hearing threshold, i.e. sensation level (SL), as a measure of loudness (Fletcher and Munson, 1933). However experimental data proved this measure to be inadequate. Since then, various formulas and methods for loudness calculation have been developed and improved by incorporating new knowledge of hearing mechanisms. In the following sections, methods developed by Fletcher's group, by Stevens, and by Zwicker and colleagues, are discussed in chronological order. Lastly, the hearing principles underlying these methods are summarized and compared to each other.

3.1 History

3.1.1 Methods Developed by Fletcher, *et al.* at Bell Laboratories

The significance of the work by this group to the field of psychoacoustics can be summarized by a quotation from Davis (Stevens and Davis, 1983) in the first paragraph (pp.xvii) of the Preface to the Paperback Edition of *Hearing: Its Psychology and Physiology*.

Hearing: Its Psychology and Physiology was written in 1936-1937 at a time when psychoacoustics had matured Modern techniques for generating and measuring sound had been developed in the 1920s and they had been applied to the description of human auditory abilities. The Bell Telephone Laboratories, particularly the group headed by Harvey Fletcher, had let the way, and their data, . . . , are still valid. Additions have been made, of course, but few modifications of the basic ideas have been necessary. – H. Davis (1983)

Fletcher and Steinberg, 1924

In an attempt to compute the effect of transmission loss on the loudness of a speech signal, Fletcher and Steinberg found it depended on the distribution of the sound energy as perceived by the ear, e.g. the spectral composition of the sound, and the hearing thresholds of the person. Furthermore, the ‘loudness loss’ from each frequency component contributing to the ‘total loudness loss’ of the speech sound was in proportion to a *fractional power* of the energy of the component. Using intensity sensation level in dB as a measure for loudness, Fletcher and Steinberg presented a formula for computing the ‘total loudness

loss', $\bar{\alpha}$, of a speech sound in dB SL, given the 'loudness loss', α at each frequency.

$$(10^{-\bar{\alpha}/10})^b = \int_0^\infty G(n)(10^{-\alpha/10})^b dn \quad (3.1)$$

The parameters $G(n)$ and b in Eq. 3.1 represented a weighting factor depending on frequency and the fractional power, respectively. Their values had to be determined empirically.

By systematically low-pass and high-pass filtering a stimulus and by asking the subjects to adjust the level of unfiltered stimulus to match the filtered stimulus in loudness, Fletcher and Steinberg were able to calculate the values of $G(n)$ and b . They found that both values depended on the stimulus level and that the nonlinearity of the ear's transmitting mechanism played an important part in determining the loudness of complex sounds. The exponent b varied between 1 and 1/3 depending on SL. For speech between 70 and 100 dB SL, b was 1/3, and $G(n)$ could be obtained by the filter slope. For a complex sound of discrete frequency components below 30 dB SL, b was 1, and $G(n)$ was found to correspond to the hearing threshold, providing a definite physical meaning. It was the fraction of the total energy in the *inner spectrum* that was carried by each component. The inner spectrum was obtained from the stimulus spectrum in the ear canal reduced by the relative sensitivity of the ear for various frequencies.

Above 30 dB SL, the fractional power decreased with increasing sensation level of speech. At 100 dB SL, the loudness of a frequency region was related to the cube root of the intensity weighted by the threshold of the region.

Important concepts laid down by this initial work were: (1) the loudness of a sound is a summation of loudness contributed by the individual component of the sound; (2) only

a fraction of the component loudness contributed to the total loudness (a prescient view of the role of masking in loudness summation); (3) frequency regions rather than individual frequency components provide a more accurate summation of loudness (a prescient introduction of the concept of a critical band); (4) below 30 dB SL, the fractional power was 1, indicating that the loudness of each frequency component was *linearly proportional* to the energy weighted by the reciprocal of the threshold of the component. This last item is similar to Hellman and Zwislocki's later proposal (1961) that, near threshold, the loudness, in sones, grows in proportion to the intensity.

Fletcher and Munson, 1933

Almost ten years later, Fletcher and Munson (1933) reviewed the works in progress for loudness calculation and presented more rigorous definitions of variables used in loudness measurements, a new formula for loudness calculation, and experimental data for deriving various parameters in the formula. Since this formula and the principles supporting it laid an important foundation for loudness calculation methods developed later by others, they are discussed here in some depth.

The variables defined were sound intensity, reference intensity, intensity level, reference tone, loudness level, and *loudness*. The definitions of these variables have remained virtually unchanged, except for some minor variations in the definition of loudness level and loudness. A 1-kHz pure tone was used as the reference stimulus for loudness comparisons. The loudness level of a test sound was defined as the intensity level of the reference tone that was equally loud to the test sound. Loudness was defined, not in dB sensation level, but as a number

" N [which] is proportional to the number of nerve impulses per second reaching the brain along all the excited nerve fibers when the typical observer listens to a simple tone." (p.85)

The definition was based on the assumption that the number N

"measures the magnitude of the auditory sensation produced when a typical individual listens to a pure tone. Since by definition the magnitude of an auditory sensation is the loudness, then N is the loudness of this simple tone." (p.85)

For a pure tone, the loudness N depended on both the frequency f and the intensity β as represented by Eq. 3.2

$$N = G(f, \beta) \quad (3.2)$$

where G was a loudness function determined by any pair of values of f and β . If f was 1 kHz, then β was the loudness level.

For a complex tone, the loudness N of the tone was the sum of fractions of loudness contributed by all frequency components. Because of the interference of neighboring frequency components, only a fraction of each component's loudness contributed to the total loudness of the complex tone. Let k indicate the k th frequency component of the complex tone, and let the loudness of the k th component be N_k , which depends upon the component's frequency f_k and intensity β_k . The equation for calculating the loudness of this complex tone is given as:

$$N = \sum_{k=1}^{k=n} b_k N_k = \sum_{k=1}^{k=n} b_k G(f_k, \beta_k) \quad (3.3)$$

where b_k is a value between 0 and 1 indicating the fractional amount of loudness N_k for the k th component contributes toward the total loudness N of the complex tone, and n is the number of components.

Iso-loudness Contours If the loudness level of every component L_k was known, then $G(f_k, \beta_k)$ could be replaced by $G(1000, L_k)$. If the relation between any possible combination of f and β with the loudness level L could be established, then the function G could be simplified to $G(1000, L_k)$. To establish this relation, an experiment was conducted to measure the binaural loudness levels for 10 pure tones at 4 to 15 levels. The loudness functions for the 10 pure tones are shown in Fig. 3.1 where the *'s (data points with the lowest intensity levels) represent the detection thresholds of the tones and o's represent the mean loudness levels. The loudness functions (curves) were obtained from a least-square fit to the data. The iso-loudness levels from 10 to 110 dB in 10 dB steps are interpolated from these fitted loudness functions to form the iso-loudness curve, shown in Fig. 3.2. The o's represent the intensity levels of the pure tones corresponding to the indicated loudness levels and the *'s represent the detection thresholds. These loudness contours deviate slightly from the Fletcher and Munson's loudness level contours because of the curve fitting used here for the loudness functions. Both the loudness level functions in Fig. 3.1 and the iso-loudness contours in Fig. 3.2 show that the loudness of pure tones varies with intensities and frequencies.

Development of Loudness Function, $G(1000, L)$ The loudness function developed here was very important in the history of loudness measurement. It was the first attempt toward developing a loudness scale that would describe the loudness growth on a linear scale

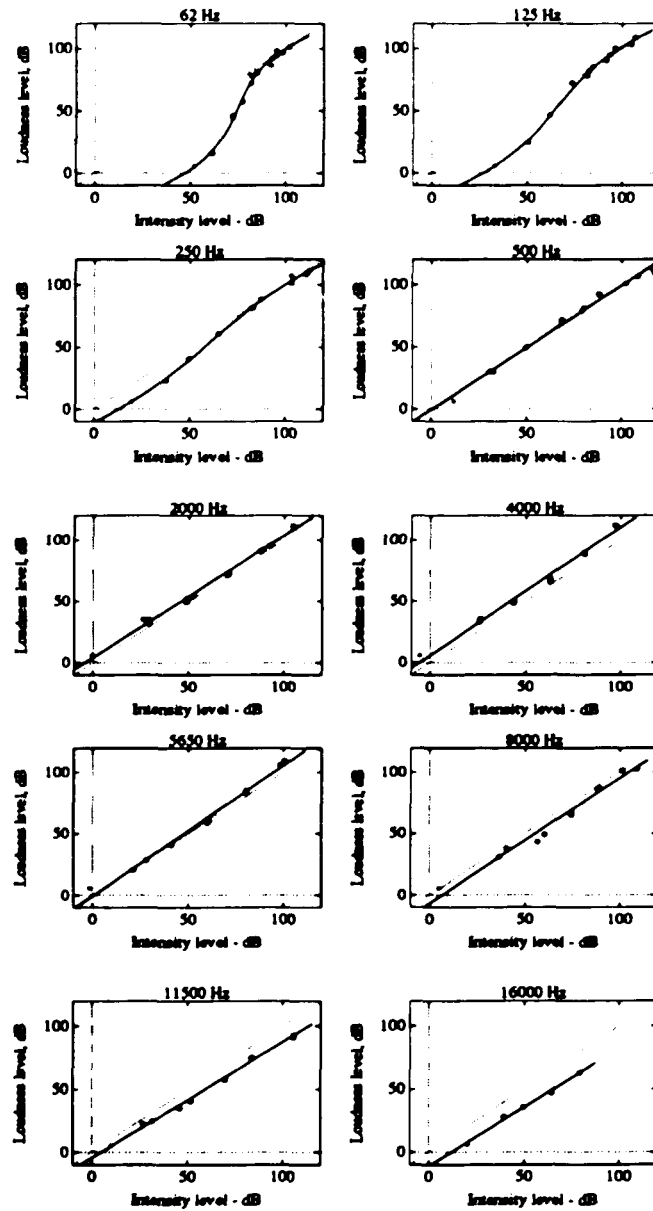


Figure 3.1: Loudness Level Data for 10 Pure Tones Measured by Fletcher and Munson, (1933). The lines representing loudness level functions were fitted to the measured loudness level data using polynomial least-square fit (by Jeng).

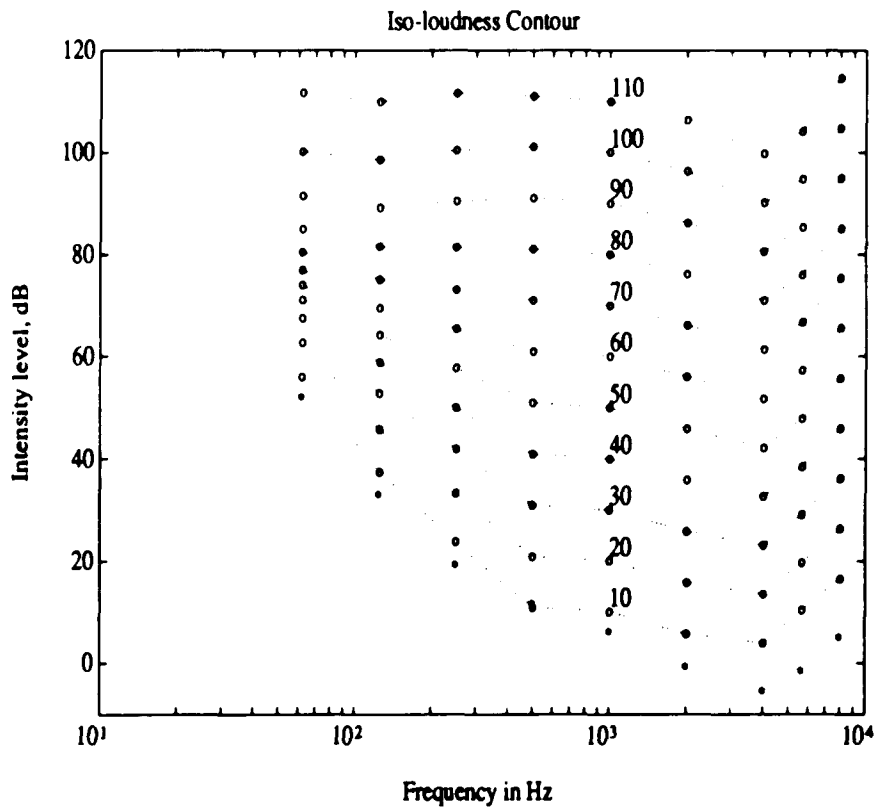


Figure 3.2: Iso-loudness Contours Derived from Loudness Levels Data Measured by Fletcher and Munson. The contours are an estimation of the iso-loudness contours derived by Fletcher and Munson (1933). The open circles along each vertical line represent the loudness levels of the corresponding frequency at 10 to 110 phons at 10 phon steps, where phon is the dB SPL of the equally loud 1-kHz reference. These loudness levels were obtained using interpolation from the loudness level functions in Fig. 3.1. The dotted lines connect the interpolated loudness levels of equal phons to provide an approximation of Fletcher and Munson iso-loudness contours. The abscissa represents frequencies of the tones and the ordinate represents the intensity levels of the tones in dB SPL. Any point, (x, y) , on the graph represents a pure tone at frequency x and at y dB SPL. The tones on the same contour sound equally loud.

and would relate the loudness (N) in a linear scale to the sound pressure level (SPL). This relation mapped the perceptual magnitude of a sound, the loudness, to its physical attribute the intensity. This loudness scale was linear. Let N_1 and N_2 represent the loudnesses of sound 1 and sound 2, respectively. If N_1 is twice as large as N_2 , then sound 1 is twice as loud as sound 2. If $N_1 = N_2$, then $N_1 + N_2 = 2N_1 = 2N_2$, provided that the two sounds did not overlap in excitation on basilar membrane. This loudness function preceded the one used today, the sone scale recommended by Stevens in 1936. The difference between the two scales is that the loudness scale developed by Fletcher *et al.* was based on hypotheses drawn from physical and physiological evidence, while the sone scale was developed based on experimental psychology and mapped to a number system with an arbitrary origin. The two curves were the same and described the same function adhering to the same mathematical rules except the numbers used to represent loudness were shifted. It is not clear to this day why the sone scale overrode the one developed by Fletcher and Munson.

The loudness function G was determined by experimental data that matched (1) the total loudness of two equally loud pure tones to the loudness of one alone, (2) the total loudness of 10 equally loud pure tones to the loudness of one tone alone, and (3) the loudness of a tone heard in one ear to that in two ears. For (1) and (2), the assumptions were that if the components of the complex tone were widely separated in frequency and had equal loudness levels so that each tone stimulated separate areas on the basilar membrane and contributed equally to the auditory sensation, then the loudness of the combined tone should be the sum of the loudnesses of the individual tones with b_k equal to 1. The formula for calculating the loudness of the complex tone in this case was specified as: $\sum_{k=1}^{k=n} b_k G(1000, L_k) = nG(1000, L_k)$. For (3), it was assumed that a sound heard

monaurally would be half as loud as the sound heard binaurally. This was based on the assumption that listeners were *able* to indicate that loudness had been halved when the magnitude of the sensation was decreased to one-half, the magnitude being the number of nerve discharges reaching the brain per second. Data collected in these three protocols were used to determine the function G with the function value for 0 loudness level set arbitrarily to unity. The rest of the function values were built relative to this point. The complete set of values of this function, for loudness levels from -10 to 129 dB in 1 dB step, were given in Table III in Fletcher and Munson's paper.

Development of the Fractional Amount, b_k For complex sounds, where components had different loudness levels and was closely packed together, the derivation of the b_k 's was much more complicated due to the masking effect from adjacent components. The fractional amount, b_k , that the loudness of one component contributed to the total loudness depended roughly on its loudness level, its relative frequency and its relative loudness level compared to other components. Since a tone is much more effective in masking another tone of higher frequency than of lower frequency, the lower frequency components were the main concerns in evaluation of the masking effects on b_k . Let f_k and L_k be the frequency and the loudness level of a component, let f_m and L_m be those of a masking component, and let Δf and ΔL be the differences in frequency and in loudness level between the k th component and the masking component.

Introduction of the Critical Band Concept In the case where Δf is very small, e.g. 10 Hz, Fletcher and Munson observed from the empirical data that it was necessary to group components within some bandwidth for multiple components to be summed

and counted as one single component. The intensity I and the frequency f of the new component should be summed as follow: $I = \sum I_k = \sum 10^{\beta_k/10}$ and $f = \sum f_k I_k / I = \sum f_k 10^{\beta_k/10} / \sum 10^{\beta_k/10}$ where β and β_k were intensity levels corresponding to I and I_k . They proposed that the bandwidth depended on the frequency of the component: for frequencies below 2 kHz, the bandwidth was 100 Hz; for frequencies between 2 and 4 kHz, the bandwidth was 200 Hz; for frequencies between 4 and 8 kHz, the bandwidth was 400 Hz; for frequencies between 8 and 16 kHz, the bandwidth was 800 Hz. These bandwidths were determined empirically to give the best results. This relatively simple loudness model appears to be the first to involve the important psychoacoustical phenomenon known as the *critical band*.

Four conditions on ΔL were considered to provide the final formula for b_k , where $\Delta L = L_k - L_m$. Firstly when $\Delta L = 0$, i.e. all components had equal loudness levels, b_k for the *fundamental frequency* component was 1, and for the rest of the components, $b_k = [(250 + \Delta f)/1000]Q(L_k)$. $Q(L_k)$ was a function of loudness level. The formula showed that b_k depended on Δf , and the larger the Δf , the less the interference between the two components, and the closer b_k was to 1. The interference between two components also depended on the loudness level of the components. Function Q , built upon the empirical data, provided a level dependent index to the interference. For convenience, the L_k were replaced by the intensity levels β_k in function Q as $Q(X)$ where $X = \beta_k + 30 \log f_k - 95$.

In the second case, when L_k fell below the masking threshold of the masker at L_m , this k th component contributed nothing to final loudness, so b_k was set to zero. Let T be the level difference between L_m and L_k such that the k th component was just masked out by the masking component. Then this case could be presented as: $b_k = 0$, when $\Delta L = -T$.

The values of T depended on the Δf . In the third case, when $\Delta L > 12$ dB, then $b_k = 1$. This is when L_k was 12 dB greater than L_m , the masking component had no effect on the k th component. In the fourth case, where $-T < \Delta L < 0$, the factor determining the effect of ΔL on b_k was $10^{\Delta L/T}$. Combining all four cases, the final formula for b_k was given as follow:

$$b_k = [(250 + f_k - f_m)/1000]10^{(L_k - L_m)/T}Q(\beta_k + 30\log f_k - 95).$$

The method was tested by comparing its calculated results with the observed results for various steady-state complex tones over a large intensity range. Most of the calculated results matched well with the measured data.

New concepts laid down by this work were: (1) mapping out the loudness levels for tones for various frequencies at various intensities using equal-loudness contours for pure tones; (2) obtaining a linear loudness scale that gives true proportion of loudness perception and a loudness function that relates loudness levels (phons, i.e. SPL of equally loud 1-kHz reference tone) to loudness number (loudness scale); (3) deriving the fractional factor of loudness for each component based on the masking effect of the neighboring component; and (4) introducing the concept of a critical band (although not referred to as such). Item 1 provided an understanding of pure tone loudness over a large range in both frequency and intensity domains.

Fletcher and Munson, 1937

Fletcher and Munson developed another method for calculating loudness based on the assumptions that the masking pattern of a complex sound represented the excitation pattern of the sound along basilar membrane (BM) and that the excitation pattern gave rise to

loudness. According to this view, the integral of the excitation per unit length along the length of the basilar membrane should predict the total loudness for the sound. Let M be the amount of threshold shift in dB for a tone heard in quiet and in noise, let x represent a place on the BM as a distance from the helicotrema, and let $F(M)$ be the excitation per unit length on the BM, dx , in response to the amount of masking M . Then the total noise loudness N can be calculated, based on Eq. 3.4, as

$$N = \int F(M) \cdot dx. \quad (3.4)$$

To solve this equation for $F(M)$, Fletcher and Munson used a wide band noise that produced a masking pattern that would have constant M covering approximately the whole length of the BM. The masking pattern and the loudness level of this noise were measured at six intensity levels from 34 to 99 dB. The integral was calculated over the distance on the BM from the place corresponding to the maximum displacement for 200 Hz to that for 1 kHz. Let x_1 and x_2 represent the two places on the BM. The mapping from frequency to position on BM was derived by Steinberg (1937). The loudness value, N , for the wide band noise at each of six intensity levels was computed using the loudness function $N = G(1000, L)$ derived from earlier work. Having determined the values for N , M , and x_1 and x_2 , the function $F(M)$ was computed as follows:

$$F(M) = N/(x_2 - x_1).$$

This equation provided the important relation between M , the amount of masking in dB, and $F(M)$, the loudness contributed per unit length of the BM. The value for $F(M)$ took into account the complicated masking effect introduced by all the neighboring components.

However, it only took care of the case when the masking levels of all components were similar and when the spectrum was continuous.

To apply the method, first one had to obtain the masking audiogram of the sound expressed in the amount of threshold shift, M , as a function of frequency. Then the frequencies were converted to positions, x , on BM and M to $F(M)$ using the conversion plot or table. The area under this masking audiogram expressed in x and $F(M)$ gave the loudness number N . Fletcher and Munson provided two charts for loudness computation to simplify the computation procedure and to make the application easier. However, one still had to measure the masking audiogram. To evaluate the method, they measured the masking audiograms for bands of noise at narrow, medium, and wide band widths at five intensity levels. The bands of noise, narrow and medium band widths, were also measured at three different center frequencies. The comparison of the loudness levels observed and calculated for these stimuli were very good except a few cases with medium band width at high intensity levels with low or high center frequencies. The test for pure tones was poorer because it was difficult to measure tone masking patterns near the tone frequencies because of beats and other tone-on-tone interference.

In order to simplify their model further, Fletcher and Munson developed a method for calculating the masking audiogram from the intensity spectrum of the sound. The calculation was based on an observation that subsequently lead to the concept of critical bands. They found that a band of noise of about 200 Hz wide had the same masking power over a pure tone as wider bands of the same spectrum levels. Based on this discovery, they derived a set of iso-masking curves that related the spectrum level, B , in dB to masking level in dB, M , as a function of frequency. This method worked well for sounds with continuous

spectrum, but did not predict masking audiograms accurately for sounds containing a small number of discrete frequency components.

New concepts laid down by this work were: (1) the use of masking patterns to estimate the excitation patterns of maskers (the sound whose loudness are to be calculated); (2) relating the frequency scale to position on the BM so that the excitation pattern can be expressed in terms of position on the BM; (3) using uniform masking pattern to estimate fractional loudness contributed by each unit length on the BM after taking into account the masking effect from all neighboring excitation.

Fletcher, 1938

In this paper, Fletcher reviewed the development of the loudness function, the development of his BM position-frequency function, several loudness calculation methods for various types of stimuli, and the relationships between loudness, masking, and the hearing process. The methods for calculating loudness for various types of stimuli have been discussed in previous sections, so only the new developments are discussed here.

The loudness function first developed by Fletcher and Munson in 1933 was revised several times after more data became available. This function was used by Steinberg and Gardner (1937) to explain two types of hearing loss, constant type (conductive) hearing loss and variable type (sensory or neural), and to calculate hearing loss.

BM position functions relating the position of maximum stimulation on the BM to the frequency of the tone stimulating it had been derived using physiological or psychophysical data by various researchers. Fletcher developed this function based on masking measurements of wide band noise. His function corresponded closely to others, however the length

unit used by him was different. Other functions used absolute length in mm to represent position on the BM, while he used percent of total auditory neurons as a measure. He assumed that along BM were patches of auditory nerve terminals, each patch of the nerve terminal corresponded to some position on BM where the maximum displacement for frequency f occurred. Based on the population density of auditory neurons along the BM computed by Guild *et al.* (1932), his function related a frequency to a specific patch of nerve terminals located at some percent of total nerves from the helicotrema. According to Guild *et al.* there were fewer terminals on the two ends of the BM and the highest number along the middle and basal turns of the cochlear.

He also presented his hypothesis of the neural coding of loudness in explicit terms, which provided consistent support to his method of loudness calculation. In light of our current understanding of auditory physiology, his hypothesis was very reasonable. However, in his day and for a long time after, his approach was not accepted by researchers in the area and his loudness scale and loudness calculation methods were lost.

Fletcher's central hypothesis stated that there were two processes involved in loudness perception. The two processes were

“At low levels the most sensitive nerve endings begin to send discharges to the brain. As the intensity increases more discharges occur from these sensitive nerve endings and also more endings are excited.” (p.293)

The quotation indicated, in today's terms of auditory nerve physiology: (1) each auditory nerve had a rate-level function, and (2) associated with each characteristic frequency were many auditory neurons with different sensitivity thresholds.

The second process has been confirmed by Liberman's experiment in 1978. The experiment compared the auditory-nerve responses from four cats that were raised in a low-noise chamber, to those of routinely available cats, to determine if the latter were affected by chronic noise exposure. One study of the experiment showed that associated with each characteristic frequency was a group of auditory nerves. The nerves in each group could be classified into three subgroups based upon their spontaneous rates, high-, mid-, and low-spontaneous rates. Those fibers with high-spontaneous rate had low sensitivity thresholds, those with low-spontaneous rate had high sensitivity thresholds. Out of all fibers of similar characteristic frequencies, 61% were high-spontaneous, low-threshold fibers, 23% mid-spontaneous, mid-thresholds fibers, and 16% low-spontaneous, high-threshold fibers. The high- and mid-spontaneous fibers appeared to have thresholds within 10-15 dB of each other. However, the thresholds of the low-spontaneous fibers spanned over a 50 dB range. This data showed that as the input stimulus (a pure tone) pressure level increases, fibers, tuned to the frequency of the stimulus would be excited as the thresholds of the fibers were exceeded.

Fletcher further proposed that

the steep part of [the loudness function (below 25 dB SPL) was] due principally
to the first process and the constant slope part [was] due to the second process.

(p.293)

The loudness scale he developed assigned 1 to the intensity at threshold and assigned 1,000,000 to the highest intensity level the ear could bear. He argued that the loudness perception was proportional to the number of active fibers, so the loudness scale, with numbers corresponded to the total number of spikes per second (rate) reaching the brain.

was a reasonable one.

New concepts incorporated in this approach were: (1) an improved version of the critical band concept; (2) an extended loudness calculation method to estimate loudness perceived by the impaired ear using its audiogram.

Fletcher, 1950

In this paper, Fletcher developed a method for calculating hearing loss for speech from an audiogram. The formula yielded a calculated value that gave better predictions than the rule in common use that was to take the average hearing loss at 500, 1000, and 2000 Hz. The prediction was especially good for steep loss. The method was based on the concept for calculating loudness developed by Fletcher and Munson (1937). The relevant part of the calculation relevant was the formal inclusion of *critical band* using critical bandwidths for loudness summation.

Summary

The preceding review reveals that much of the theoretical work and fundamental concepts underlying loudness summation of complex sounds used were introduced by the Fletcher and Munson group. Their formulations of these concepts were based upon the limited physiological knowledge available at that time, long before many of the basic discoveries in auditory physiology. Yet these assumptions have proven to be quite accurate. The works on loudness summation also brought about discovery of many important psychophysical phenomena, such as critical bands in masking, and loudness summation.

The pioneering work of the Fletcher and Munson group has become the foundation for other methods of calculating the loudness of complex sounds, including Stevens' method

and Zwicker's method which are described below.

3.1.2 Stevens' Method

Stevens, 1956

Based upon the proposal by Gates (1937), and work elaborated on Gates' proposal by Beranek *et al.* (1951), Peterson and Beranek, (1954), and by Mintz and Tyzzer (1952)¹, Stevens presented a method for calculating the loudness of a complex sound. The original proposal was that given the octave-band spectrum of a complex sound, its loudness could be computed by summing the loudness of individual octave bands. Stevens argued that the sum of the loudness of individual octave bands was not the same as the loudness of the sound with all bands present due to *mutual inhibition* between bands. He differentiated the mutual inhibition from masking, in that the term masking was used to indicate the threshold shift of a tone in quiet and in noise. Mutual inhibition was to mean the reduction or suppression of loudness.

Stevens' method of loudness calculation was basically as follows: given a complex noise represented by an octave-band spectrum with n octave bands, let the loudness, S , of each band be equal, i.e. all bands are equally loud. Let R be the reduction of loudness in each band. Then the total loudness, S_t (in sones), of the complex noise could be computed by the following formula:

$$S_t = S + (n - 1)(S - R). \quad (3.5)$$

The Eq. 3.5 indicates that when there is only one band in a complex sound, the loudness of the sound is the same as the loudness of the band. So, the first S is not reduced. When

¹These works were based on the work by Fletcher and his coworkers.

a second band is added to the sound, the loudness contributed by the second band to the total loudness is reduced by R . So the total loudness S_t is $S + (S - R)$. When a third band is added to the sound, the loudness contributed by the third band to the total loudness is also reduced by R . Then the total loudness is $S + (S - R) + (S - R)$, which is $S + (3 - 1)(S - R)$.

Stevens suggested that if the loudness of all bands were equal then the reduction of total loudness for each additional band would be the same. In consequence, the fraction of loudness contributed by each additional band should be the same. Let F represent the fraction of the loudness of a band, such that $F = (S - R)/S$. Then Eq. 3.5 could be replaced by $S_t = S + F(\sum S - S)$. However, noise did not usually come with equal loudness in all bands. A first approximation was given in the following equation:

$$S_t = S_m + F(\sum S - S_m) \quad (3.6)$$

where S_m was the loudness in sones for the band with most energy.

To complete the calculation, (1) the loudness function that related loudness levels (phons) to loudness in sones (the loudness scale developed by Stevens) had to be formulated; and (2) the fraction F for various conditions had to be derived empirically. The loudness function in sones could be calculated using Eq. 3.7:

$$\log_{10} S = 0.03P - 1.2 \quad (3.7)$$

where S was loudness in sones and P was loudness levels in phons. A table relating phons to sones was also provided in Appendix I of Stevens (1956).

To derive F , experiments were conducted to measure the loudness (sones) of individual

octave bands from 75 Hz to 9600 Hz, i.e. 75 to 150 Hz band, 150-300 Hz band, etc., and the loudness of two octave bands, of three octave bands, of four octave bands, of six octave bands, and of seven octave bands. (Not all combinations were measured.) The data were used to give a scatter plot of $\sum S/S_m$ vs. S_t/S_m . The slope of the line fitted to the plotted data provided the value for F which was 0.27 and was rounded up to 0.3. This indicated that only 27% of the loudness of each additional band contributed to the total loudness. However the data showed a slight dependency on loudness level such that at low and high ends of the range F values were lower, about 0.13, and at mid-range the F values were higher, about 0.38. After examining all possible errors that might contribute to the derivation of F value, Stevens suggested that the variance of F due to the loudness level was negligible. The final equation for calculating the total loudness for a steady noise with a continuous spectrum expressed in octave bands was:

$$S_t = S_m + 0.3(\sum S - S_m). \quad (3.8)$$

To simplify the calculation, Stevens conducted experiments to obtain a set of four equal-loudness contours for octave bands in diffused sound field. The loudness levels of the contours ranged from 45 to 97 phons. By interpolation and extrapolation, he then derived a set of loudness functions for octave bands to relate the sound pressure levels (SPL) of each octave band to loudness in sones. This set of loudness functions was also represented in a set of nomograms that provided sone values for corresponding SPL's in all octave bands. Having obtained the sone values, Eq. 3.8 could be used to compute the total loudness, S_t . Finally, another nomogram related sones to phons, the loudness level of the noise.

The method was used to evaluate the loudness data measured by Quietzsh (1955) for a variety of noise: automobile horns, airplanes, trucks, trains, motorcycles, milling machines, air hammers, doorbells, typewriters, etc. The prediction results in general agreed with the measured data with some large deviations up to 10 phons off. The poor results were for noise of line spectra, recorded from original sources, and of impulsive noise. The loudness levels of the tested noise were from 50 to 110 phons. The discrepancies, according to Stevens, were in part due to the meters used to measure the SPL's of the noise. Comparisons for hypothetical (generated) white noise were very good.

Eq. 3.8 was also extended to calculate loudness levels for noise represented in a third-octave band spectra and in a half-octave band spectra. The only difference in the formulas was the F value. For the third-octave spectrum, the F value was 0.13 and for the half-octave spectrum, the F value was 0.2. The loudness functions relating SPL of the third-octave and half-octave bands to loudness in sones were interpolated from those for octave bands.

The differences between this method and that of Fletcher and Munson were (1) Stevens used octave band spectra which were easily measurable because of the improvement in spectrum analyzers after more than 20 years; (2) loudness functions were measured in sones (as defined by Stevens); (3) and Stevens' basic philosophy was to not consider the physical correlate of the psychophysical measurements. The last item was directly contrary to Fletcher and Munson's approach, that is they based all their hypotheses on possible physiological correlates of the hearing process.

Stevens, 1961

In 1960, the International Standard Organization (ISO) requested provision of a numerical table of equal-loudness contours for octave bands as part of a requested amendment of Stevens' method before adopting the method as an ISO standard. In response to the request, this revision was carried out. The changes were in two areas: (1) the equal-loudness contours were replaced by approximations using straight line segments, and (2) the loudness function relating phons to sones was modified to reflect the loudness growth pattern of a wide band noise. This version of the method was referred to as Mark VI (Stevens, 1961).

Each equal-loudness curve was approximated by three line segments: the first line segment over low frequency region had a slope of -6 dB/octave; the second line segment over mid-frequency region had a slope of -3 dB/octave; and the third line segment over high frequency region had a slope of 12 dB/octave. The joints between the first and the second line segments fell on frequency points depending on the band pressure levels of the contours. These frequency joints could be approximated by a straight line which intersected with 1 kHz at a 10 dB band pressure level and with a -21 dB/octave slope. The joint frequencies between the second and the third line segments fell on 9 kHz independent of band pressure level. Since the equal-loudness contours were replaced by approximations, the term 'loudness', in sones, used to indicate the loudness associated with the contours was then replaced by the term 'loudness indices'. Hence, the two terms, S_m and S , on the righthand side of the loudness summation formula, $S_t = S_m + F(\sum S - S_m)$, are defined to be the loudness indices, rather than loudness, of the corresponding bands. The term S_t remains the total loudness in sone.

In the original publication of Stevens' method (1956), it was observed that the F value was loudness-level dependent, highest in mid-level range, and lower on both low and high level ranges. For convenience of the calculation, F value was fixed at one value. Later, critical band data also demonstrated that loudness grew faster over the mid-level region for a band of noise wider than the corresponding critical bandwidth. In order to compensate the error caused by keeping the F value fixed, the function used to relate sound pressure levels of the band to the loudness index was revised to reflect the so-called *midlevel bulge* from 20 to 60 dB. The midlevel bulge replaced the simple power function shown in Eq. 3.7 which could be represented by a straight line.

This version of Stevens' method of loudness calculation was later adopted nationally as an American National Standard (ANSI S3.4/1968) and internationally as (ISO/R 532-1966(E)).

Stevens, 1971

The procedure, Mark VII, was an improved version for calculating *the perceived level of loudness or noisiness* from Mark VI, according to Stevens, in the sense of bringing the theory supporting the procedure to closer alignment with the performance of the ear and the auditory nervous system. The basic procedure was the same as that for Mark VI. The differences between Mark VII and Mark VI are described as follow. Firstly, the loudness level used to indicate the loudness of a sound was changed to the *perceived level in dB (PLdB)* because the procedure Mark VI was intended to predict both the loudness and the noisiness of the sound. Noisiness of a noise was also referred to as the annoyance or acceptability of the sound at the time. The term PLdB and the procedure for calculating

the noisiness of a noise were first established by Kryter (1959) based upon Stevens' Mark II (1957a) which was a simplified version of Stevens' original method (1956).

Secondly, the set of equal-loudness contours were replaced by a set of frequency-weighting contours that was derived by taking the means and medians of 25 sets of experimental data for equal-loudness, equal-noisiness, equal-annoyance, or equal-acceptability contours using bands of noise. The variety of bandwidths included 250-mel bands, octave bands, critical bands, third-octave bands, and tenth-octave bands. The frequency-weighting contours were composed of 5 or 6 line segments instead of 3 as in Mark VI.

Thirdly, the third-octave band centered at 3150 Hz, instead of a 1-kHz pure tone, was defined as the standard reference sound. Stevens argued that the growth of the apparent magnitude (loudness, annoyance, or noisiness) of the 1-kHz pure tone was not a simple power function of sound pressure, especially at low pressure levels, while that at 3150 Hz was closer to a simple power function. He believed that the power function was basic to the functioning of the auditory system because it was in all the other sense modalities. In order to keep the same sone scale for the perceived magnitude (in place of loudness), the sound pressure level of the new reference equal to 1 sone was set at 32 dB SPL which is approximately equally loud to the old reference at 40 dB SPL. The formula for calculating the perceived magnitude, S , was $S = k(E - E_0)^{1/3}$, where E and E_0 correspond to the power of the band pressure level of the reference and that of the reference at threshold. The power E was proportional to the square of the sound pressure. The threshold of the reference was -3 dB re 20 μ Pascal. (20 μ P is 0.0206 dB SPL.)

The exponent 1/3 was based upon the recommendation of Fletcher (1935) and Knauss (1937). They indicated that above 40 dB SPL, the loudness function for 1-kHz tone could

be represented by a straight line that related loudness to the cube root of intensity. Below 40 dB SPL, loudness was directly proportional to intensity. Knauss proposed the equation $L = I(10^{-5/2}I + 1)^{-2/3}$ to represent the loudness function, where L was the loudness in millisones and I was the intensity in units of $10^{-16} \text{ watts/cm}^2$. Furthermore, extensive cross-modality experiments (Stevens, 1966; Moskowitz, 1968; Stevens and Greenbaum, 1966; Stevens, J. C. and Marks, 1965; Stevens, J. C. and Stevens, S. S., 1963) had confirmed this exponent value, 1/3 for intensity and 2/3 for pressure, for the loudness function.

Fourthly, the F value used in the summation rule, $S_t = S_m + F(\sum S - S_m)$, was changed from a fixed value to a function of the perceived magnitude (sones). The change was based on the observations of loudness functions for noise and of equal loudness contours for bands of noise with varying bandwidths (critical band experiments by Zwicker, Flottorp, and Stevens, 1957; Scharf, 1959). Lastly, when the total loudness, S_t , in sones was obtained, the corresponding perceived level, P , in PLdB could be calculated using formula: $P = 32 + 9 \log_2 S_t$. The value 9 corresponded to the dB increase per doubling of loudness.

In this paper, Stevens also reported on the outcome of the loudness predictions by Mark VI and by Zwicker's method for a large variety of noises by various groups. It was found that the mean difference between the measured loudness level and the predicted loudness level by Mark VI was -0.4 phon, while the mean difference between the measured and the predicted loudness levels by Zwicker's method (assumed to be 1968 edition of ISO 532 B or around that time) was 2.3 phons.

3.1.3 Methods Developed by Zwicker *et al.*

Zwicker and Feldtkeller, 1955

Zwicker and Feldtkeller computed the loudness of a sound from its power spectrum by the following steps. Firstly, the frequency axis of the spectrum was divided into 24 consecutive *frequency-groups* whose bandwidths corresponded to the respective critical bandwidths. Secondly, the loudness for the sound intensity within each frequency-group was estimated. Thirdly, the loudness within each frequency-group was reduced, depending on the loudness of the neighboring frequency-groups. Lastly, the resulting loudnesses of all frequency-groups were summed to give the final loudness of the sound which could thus be converted to the loudness level in phons.

In the first step, a frequency-group represented a unit length for loudness summation. The frequency components of the spectrum within the frequency-group acted as one component to contribute to loudness. The band widths of frequency groups at various center frequencies were measured using equal-loudness contours as functions of varying bandwidths at loudness levels ranging from 20 to 80 dB effective sound pressure level. The data showed that by keeping the total sound pressure level constant, the loudness of the band of noise remained the same until the bandwidth was greater than that of the respective frequency group. The data also showed that the bandwidth seemed to be independent of the loudness level except for the frequency group centered at 500 Hz.

In the second step, the loudness produced by each frequency-group alone was evaluated. Firstly, the total sound pressure level of each frequency-group was converted to the loudness level using equal-loudness contours. Secondly, the loudness level was converted to loudness in sones using the loudness function. The equal-loudness contours were measured using

Békésy's tracking method. The loudness function was derived by fractionation, i.e. to half or double the loudness of a sound in comparison to the loudness of a fixed reference tone. The derived loudness function was steeper than Stevens' loudness function at low intensity levels (according to Stevens, Addendum, 1956). The loudness of each frequency-group did not take the interference from the neighboring frequency-group into account.

In the third step, loudness reduction within each frequency-group due to the masking effect of the two neighboring frequency-groups (one above and one below) were estimated using trapezoidal patterns. The masking pattern of a noise band with bandwidth equal to one frequency-group was equated to a trapezoid. The upper side of the trapezoid was equal to one bandwidth and the base equal to three bandwidths. The two flanking sides spread to the neighboring bands. The excitation patterns for two adjacent frequency-groups were represented by two overlapping trapezoids. The total loudness of the two bands was equal to the total area of the two overlapped trapezoids. Because of the overlap, the total area of the two trapezoids was not the same as the two non-overlapping trapezoids. The concept was the same as Fletcher and Munson's approach or the same as Stevens, if viewed the other way. That is, because of the overlap, the loudness contributed by the second frequency-group was reduced. The amount of reduction was level dependent.

In the last step, the total loudness of the sound was computed by summing the reduced loudness from all frequency groups. This is the same approach as the previous method. The method was assisted using graphic charts.

Zwicker and Scharf, 1965

Zwicker and Scharf (1965) summarized the psychophysical and physiological principles used in the loudness calculation methods developed and improved since the original study by Zwicker and Feldtkeller (1955). They then developed a model based on these principles. The model was evaluated against several set of data and was used to explain some observations that seemed contradictory to conventional understanding. This model has been adopted as an International standard, ISO 532b (1975).

The model assumed that the loudness of a sound was proportional to the excitation pattern at the organ of Corti produced by the basilar membrane displacement caused by the sound. When the amplitude of the sound was increased, the excitation pattern would increase in amplitude and spread over frequency, and hence the loudness increases. However, the excitation patterns did not change for bands of noise centered at the same frequency, whose bandwidths were less than the corresponding critical bandwidth, and whose energies were constant. The spread of energy has no effect on loudness until the critical band is exceeded. So critical bands were used as the integration unit for loudness summation. The human audible range was divided into 24 consecutive discrete critical bands for center frequencies from 25 Hz to 12.5 kHz. The critical bands were called frequency-groups in the original paper and the band units were referred to as *Barks*. The excitation patterns for each critical band of noise at various sound pressure levels were estimated from its masking patterns such that the excitation level was 3 dB above the masked level for low and mid center frequencies and 6 dB above the masked level for high center frequencies. To convert the excitation level in each band to loudness per band, a formula derived based on Steven's law (Stevens, 1957b) and a loudness matching

experiment were used to obtain the exponent of the power law that related the excitation level to loudness in each band. Stevens' law² used was $\Delta N/N = k'\Delta I/I$, where I is the intensity of a tone and N is the loudness. The exponent of the power function for each critical band of noise was 0.23. The basic concepts used here were, in general, not new.

The model consisted of four stages for transforming a sound to loudness in sones. The first stage converted the power spectrum of the sound to critical-band power spectrum in Bark scale. The frequency components of the spectrum were grouped into 24 consecutive critical bands (Barks) and the total SPL for each band was computed. The second stage converted each band of the spectrum in Bark to an excitation pattern. The pattern was derived from the estimated masking pattern for each critical band of noise at the corresponding center frequency and corresponding level. The model assumed that a band was not effective in masking the bands of lower frequencies and was very effective in masking bands of higher frequencies. Thus, the excitation pattern of a band did not spread toward lower-frequency bands, but had a spread toward higher-frequency bands. The amount of spread was level dependent and increased nonlinearly. The model also assumed that only the excitation above the hearing threshold contributed to loudness, so the excitation level for each band had to be adjusted in reference to the hearing threshold of the band. The third stage converted the excitation level of each band of the excitation pattern to loudness per band, defined as *specific loudness* using the power law with 0.23 as the exponent. The fourth stage summed the total loudness of the sound by integrating the specific loudness across 24 Barks.

If the sound for loudness calculation was a pure tone or a band of noise with bandwidth

²It was actually Weber's law.

less than the corresponding critical band, then there was only one excitation pattern. The integration for this should be simple. If the sound was a wide-band complex sound then there would be many overlapping excitation patterns corresponding to various Bark units. The model assumed that masking occurred wherever the excitations were overlapped. Then the maximal parts of the two excitations would be used for integration to give final loudness. The reduction of the loudness of a band due to the mutual masking of the neighboring band was implied in not counting the overlapped excitation patterns more than once. That is if the specific loudness of a pattern was less than that of an overlapping pattern, then the loudness of the lesser pattern was not counted.

The physical realization of the model was implemented through various psychophysical measurements to estimate parameters in the model.

Zwicker, 1977

Zwicker (1977) discussed five effects that could affect the loudness of time-varying sounds and extended his and Scharf's previous loudness model (Zwicker and Scharf, 1965) to account for various issues occurring in temporal integration of loudness of time-varying stimuli. A loudness meter was constructed to provide running loudness values of the sound and its temporally integrated loudness. The loudness meter was evaluated by comparing its reading with the subjectively measured data for seven types of stimuli.

Five effects on the loudness of temporally varying sounds included the phase effect, the influence of physiological noise, the effect of amplitude modulation and of frequency modulation, and the effect of high center frequency for narrow-band noise. The phase effect on loudness appeared only at low frequencies. Experiments showed that loudness

variations due to phase variation were not more than +2 or -2 dB in loudness level, not a significant effect. The internal physiological noise affected mainly loudness perception at low frequencies below 500 Hz as demonstrated by hearing thresholds and equal-loudness contours. Therefore, equal-loudness contours were imitated to account for the effect. The effect of amplitude modulation on loudness was resolved in two ways depending on the modulation frequency. For modulation frequency below 10 Hz, the peaks of the sound pressure envelope were responsible for loudness, and for modulation frequency above 10 Hz, the averaged sound pressure level was responsible. The effect of frequency modulation on loudness also demonstrated two quasi-steady states depending on the modulation frequency. For frequency-modulation frequency below 30 Hz, the loudness of the sound was perceived in the same manner as for a steady pure tone. For modulation frequency above 30 Hz, the loudness of modulated sound was related to the spectral components instead of temporal variation. Concerning the last effect, it was commonly believed that bands of noise centered at the same frequency, with the same total sound pressure levels, should sound equally loud when their bandwidths were no greater than the corresponding critical bandwidth. Contrary to this, Zwicker found that a band of noise at 5 kHz, with 1 kHz bandwidth (the critical bandwidth approximated for the band centered at 5 kHz by Zwicker) sounded louder than a 5 kHz tone at the same SPL. He attributed the discrepancy to some kind of time-dependent nonlinearities due to rapid amplitude variations, only possible at high frequencies.

The loudness calculation method for a time-varying sound was expanded based on the loudness calculation method for the steady sound, by taking temporal masking effects into account. To calculate the time-varying sound, simultaneous masking patterns were used

to estimate the spatial excitation pattern for each of 24 consecutive critical bands. Then the temporal excitation pattern for each band was estimated based on temporal masking patterns. Temporal masking included backward masking (premasking) and forward masking (postmasking). The time constant for backward masking was much shorter and less effective than forward masking, so backward masking was simulated using a fast attack time, 2 msec, and was largely ignored. Forward masking was used to estimate the temporal excitation pattern in the temporal method. It was found that the time constant and the amount of forward masking depended on the duration of the sound burst, as if there were a "critical duration" of 20 msec that divided a burst into short vs. long. A short burst gave a short forward masking pattern and a long burst gave a long forward masking pattern. The specific loudness in each band was then computed using a power law. In the next step, the specific loudness was summed across 24 bands to give the "intermediate loudness". The last step, the method applied the forward masking rules to integrate the loudness and estimate the temporal loudness pattern for the sound. The peak values of the temporal loudness pattern closely approximate the loudness of sounds such as impulses, speech, and other time-varying stimuli.

3.2 Discussion

The following summarizes various concepts used in the methods discussed in this chapter and discusses the problems associated with those approaches.

(1) Fletcher *et al.* started the search for a loudness calculation method in simple physical measurement for practical purposes but quickly moved toward psychophysics and physiology for a better understanding of the auditory system. Stevens' approach was based on

pure psychophysics, for the purpose of practical application. Zwicker *et al.* discussed their methods based on physiological principles, but implemented the methods entirely based on psychophysical measurements. These methods have not provided a full understanding of human auditory physiology, such as what the bandwidths of the human auditory filters are, how the loudness is coded by the auditory nerve fibers, and how the interactions of neighboring frequency components affect each other's firing patterns, etc.

(2) Each method used some kind of spatial unit along the basilar membrane, across which loudness was summed to give the total loudness. Fletcher *et al.* used single frequency components, grouped frequency components, critical bands, and 1% of auditory nerve patches on the BM in various methods. Stevens used one-octave bands, half-octave bands, and third-octave bands. Zwicker and colleagues used critical bands. The critical band is a psychophysical phenomenon and its physiological origin is not clear. Some of the conclusions from critical band measurement using loudness of bands of noise, or complex tones, (Zwicker *et al.*, 1957; Scharf, 1959) were questionable: (a) As the bandwidth of a noise band was reduced from critical bandwidth to the tone at the center frequency, the loudness of the bands of noise did not necessary remain the same with constant sound pressure level; (b) Was there really a clear breakpoint to indicate where the bandwidth of a critical band? (c) Were critical bandwidths truly independent of the sound pressure levels as claimed? Furthermore, the bandwidths of critical ratios as measured by Fletcher (1940) were smaller than those of critical bands as measured by Zwicker *et al.* (1957). Which one is correct? Perhaps the bandwidths measured by different methods describe different facets of the cochlear filters.

(3) The estimation of the excitation patterns using masking patterns was not completely accurate. Both Fletcher (1937, 1938) and Zwicker (1965 and 1977) used masking patterns to estimate excitation patterns for loudness calculation. Delgutte (1989, 1990) showed that the physiological mechanisms underlying pure tone masking involved at least two mechanisms, excitatory and suppressive. The suppressive element due to two-tone suppression is responsible for the rapid upward spread of masking with increasing masker amplitude. In addition, he also demonstrated quantitatively that off-frequency listening lowered masking thresholds especially for test tones around the masker. In order to derive correct excitation patterns, the effects from both the suppressive element and off-frequency listening have to be removed from masking patterns.

(4) All methods tried to take into account the effect of mutual masking between two adjacent spatial units, frequency components or bands. But masking is a complicated phenomenon. It depends on the frequency and intensity level of the masker, on the relative distance between the masker and maskee in frequency and in intensity level, and on the nonlinear additivity of masking effects from the neighboring maskers. Pure brute force was not enough to cover all possible combinations of maskers and maskees as demonstrated by Stevens in attempting to find better loudness reduction fractions. In Zwicker's method, the excitation pattern for a complex sound was formed by taking the envelope of all the excitation patterns for the 24 critical bands of the complex sound. In Fletcher's method, loudness per unit length of the basilar membrane (BM) was calculated by dividing the uniformly excited masking pattern for a wide band noise by the number of units between specific places on BM. His purpose was to take into account the complicated mutual masking between all the neighboring components. However, this only takes care of components

of fixed spectral level relation, not all combinations. True understanding of the complete physiological mechanisms underlying masking and of the relation between masking and loudness is necessary to make further progress in loudness calculation.

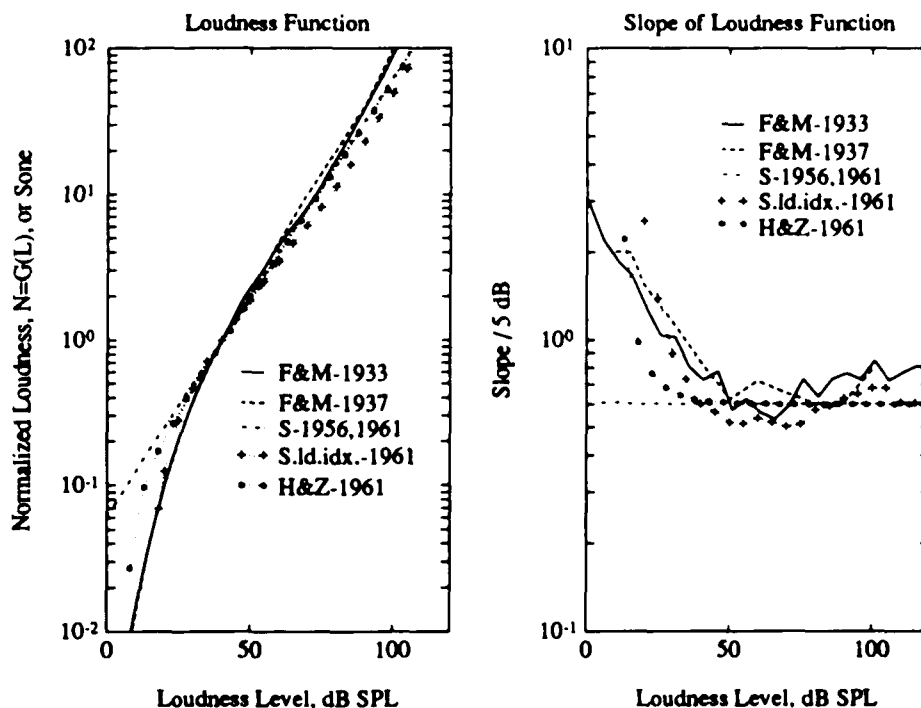


Figure 3.3: Loudness Functions for 1-kHz Pure Tone and Their Slopes. In Panel (a), loudness functions of 1-kHz pure-tone obtained using various experimental methods are shown. In Panel (b), the slopes of these loudness functions computed for each 5 dB segment are shown. The two loudness functions derived by Fletcher and Munson were used in their methods of loudness calculation in 1933 and 1937. (Table III, 1933; Fig. 5, 1937) The curve marked 'S-1956,1961' is the loudness function used by Stevens for both Mark II (1956) and Mark VI (1961). The curve is computed from Eq. 3.7, which can be expressed as the power function: $L = kP^{0.6}$, where L is the loudness in sones, P the sound pressure in μPa , and k the constant, 0.01 (Table VII, 1956). The curve marked 'S.ld.idx.-1961' is the "loudness index" function at 1 kHz, which was used in Mark VI to relate the estimated loudness of a third-, half-, or one-octave band, to the band pressure level (Table I, 1961). The 1-kHz loudness function measured by Hellman and Zwislöcki (1961) is estimated by the equation: $L = k(P - P_0)^{0.6}$, where L represents the loudness in sones, k is 0.01, P is the sound pressure in μPa , and P_0 is 45 μPa , the sound pressure at threshold. (pp.192, Carterett and Friedman, 1978)

(5) Both Fletcher's group and Stevens in their methods developed *loudness scales* for referencing the loudness of a sound.

- The first scale is the currently known *phon scale*, which is also referred to as the loudness level of a sound. The phon scale uses the sound pressure level (SPL)³ of an equally loud 1-kHz pure tone to indicate the loudness level of another sound. When a sound is judged to be equally loud as a 1-kHz tone at 50 dB SPL, the loudness level of the sound is said to be 50 phons. A function relating the sound pressure levels of a sound to its phon values is defined to be the loudness level function of the sound.

Fletcher and Munson (1933) used dB SPL of the 1-kHz tone as the loudness level and measured loudness levels for 10 pure tones from 62 Hz to 16 kHz. These loudness level functions were then used to derive the equal loudness contours. Stevens used dB SL as the loudness levels and measured loudness levels of 1-, 1/2- and 1/3-octave band noise to obtain equal loudness contours of those narrow-band noise.

- The second scale is the currently known 'sone scale', which assigns 'loudness values' to sounds, such that the loudness value is linearly proportional to the loudness percept. If the loudness of sound A is perceived to be twice as loud as sound B, then the loudness value of sound A should be twice of the loudness value of sound B. If sound A and sound B are equally loud, and the two sounds do not interact with each other, e.g. two pure tones with frequencies far apart, then the loudness of sound A plus sound B should be the sum of the loudness values of the two sounds. The *loudness function* shows the loudness of a sound as a function of the sound's dB SPL.

Fletcher and Munson (1933) defined the loudness to be the magnitude of auditory sensation, which was linearly proportional to the total number of spikes per second arriving at the brain. Based on this definition, they used three equal-loudness

³Currently, the phon value refers to dB sensation level of the 1-kHz pure tone.

measurements: binaural vs. monaural, two-tone complex, and ten-tone complex, to derive the loudness function that related the total spike rate to the sound pressure levels of 1-kHz pure tone. Their loudness functions, normalized by the loudness values at 40 dB SPL, are shown in panel (a) of Fig. 3.3, and the slopes of the function are shown in panel (b).

Stevens assumed that people, in general, were able to associate sensations caused by physical stimuli, e.g. intensities of sounds, brightness of lights, and lengths of lines, to numbers in the real number system. As the number system is linear, so should be the perceived loudness associated with the numbers. Stevens and many others used doubling, halving, loudness estimation, and loudness production to demonstrate the point. The loudness function, derived by Stevens, expressed the loudness values of a sound in *sones*⁴ as a function of dB SPL of the sound. Stevens used a 1-kHz pure tone at 40 dB SPL as a reference condition in the sone scale, and defined the reference to be 1 sone. He observed that the loudness values doubled for every 10-dB increase in sound pressure, hence he used $L = kP^{0.6}$ to describe the loudness function of a 1-kHz tone. The function and the slope of the function are shown in Fig. 3.3. In his 1961 method, a set of numbers called 'loudness index' was used to indicate the loudness of 1-, 1/2-, and 1/3-octave band noise. The function relating the loudness index to dB SPL of bands centered at 1-kHz is also shown in Fig. 3.3 (denoted by S.l.d.idx.-1961). In panel (a) of Fig. 3.3, a comparison between the loudness functions of 1-kHz pure tone by Fletcher and Munson (1933, 1937) and by Stevens (1956, 1961) shows that these functions are very similar between 35 and 85 dB SPL. Above 85 dB, Fletcher's

⁴The term, *sone*, was proposed by Stevens to name the loudness scale (footnote 10, 1936).

functions became slightly steeper. Below 35 dB SPL, Stevens' loudness functions (1956, 1961) continue to decrease at the same rate as above 35 dB SPL, while those of Fletcher's decrease at an increasing rate. The slopes of these functions can be easily observed in panel (b). The loudness index function has a shallower slope above 40 dB SPL and similar slope as those of Fletcher's below. Hellman and Zwislocki (1961), after removing many bias factors from their measurements, obtained a 1-kHz loudness function that showed steeper slope near threshold. Their function, plotted in panel (a), is estimated using the formula: $L = k(P - P_0)^{0.6}$, where L represents the loudness in sones, k is 0.01, P is the sound pressure in μPa , and P_0 is 45 μPa , the sound pressure at threshold.

The slopes of all these functions are plotted in panel (b). A comparison shows that the loudness function obtained by Fletcher and Munson as early as 1933 had demonstrated the basic property of loudness growth and was as valid as loudness functions measured by other methods.

(6) Neural coding of loudness was addressed only by Fletcher et al. in their methods. The loudness of a sound was defined by the total number of nerve impulses per second reaching the brain along all the excited nerve fibers. They further conjectured that at low sound pressure levels the most sensitive nerve endings first began to send discharges to the brain. This process was principally responsible for the steep part of the loudness function (below 25 dB SPL). As the intensity increased more discharges occurred from these sensitive nerve endings and also more endings were excited. This second process was responsible for the constant slope part of the loudness function. The same coding was proposed by Wever (1949) based on volley theory. Another neural coding hypothesis for

loudness (Whitfield, 1978) stated that loudness depended on only the number of fibers activated and the discharge rate was not a key factor in the coding of loudness. Based upon Liberman's finding (1978) of three groups of auditory nerves of similar characteristic frequency, some (Viemeister, 1988; Delgutte, 1986) has argued that the dynamic ranges of intensity levels measured at single auditory fibers are all too small to represent the total dynamic range of human hearing. When the dynamic ranges of all fibers with same characteristic frequency are combined, a larger dynamic range may be achieved

(7) All the methods for calculating loudness discussed in this section are for stationary signals, except the one by Zwicker in 1977. It has been shown that loudness growth with signal duration is nonlinear. So are the forward masking of a strong peak over following peaks and the overall perceived loudness of a temporally varying signal.

Chapter 4

EIH Based Loudness Model

In this chapter, a loudness model based on physiological principles is presented. The loudness model consists of two parts. The first part is an auditory model that simulates basic physiological functions in the peripheral auditory pathway and extracts temporal synchrony information as a representation of part of the central auditory system. The model is called the *Ensemble Interval Histogram* model, abbreviated as EIH. The second part is a loudness module that predicts the “instantaneous” loudness from the EIH output, as a function of time. The instantaneous loudness is used in deriving the overall loudness, over time, of the sound. Since the loudness model is based on the EIH model, it is referred to as the *EIH based loudness model*, which is abbreviated as EIH-LM. The structure and the underlying principles of the EIH model are described in the first section of this chapter. The assumptions and the hypotheses used to define the instantaneous loudness are discussed in the second section.

4.1 EIH Model

The physiological units of the auditory system, whose functions are simulated by the EIH model, are depicted by the block diagram in the upper half of Fig. 4.1. The units include

the middle ear (ME), basilar membrane (BM), inner hair cells (IHC), auditory nerve fibers, and the central nervous system. The parts of the EIH model simulating those physiological units are summarized by the block diagram in the lower half of Fig. 4.1 and are described in detail below.

Auditory System:



Schematics of EIH Model

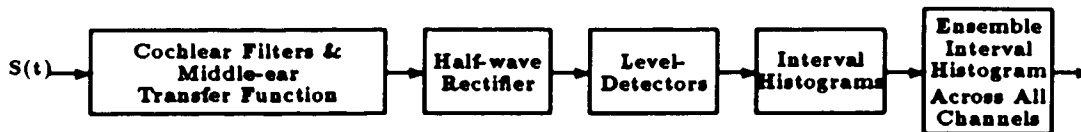


Figure 4.1: Block Diagram Depicting Functional Units of the Auditory System and Corresponding Elements of the EIH Model.

As mentioned in Chapter 2, human physiological data are not readily available, and since the peripheral auditory systems of most mammals are similar, the EIH model is based on animal data, mostly from the cat.

Cochlear Filters The middle ear transfer function and the frequency analysis ability of the basilar membrane and organ of Corti are simulated by a set of overlapping cochlear filters. Fifty-six cochlear filters are shown in panel (A) of Fig. 4.2. The extended set contains 166 cochlear filters¹. The characteristic frequency (CF) of the lowest filter is 176

¹The original set of 56 cochlear filters was provided by Delgutte from Eaton Peabody Laboratories. Each filter was derived by averaging and smoothing over several neighboring tuning curves which were obtained from three cats by C. Liberman. Ghitza tripled the set to 166 filters by interpolating additional filters. For example, filter with a characteristic frequency (CF) of x Hz was used to create two other filters, one with $CF = x - \Delta$ and the second with $CF = x + \Delta$, where $\Delta = 1/3$ of frequency difference (on a log scale) between two adjacent filters.

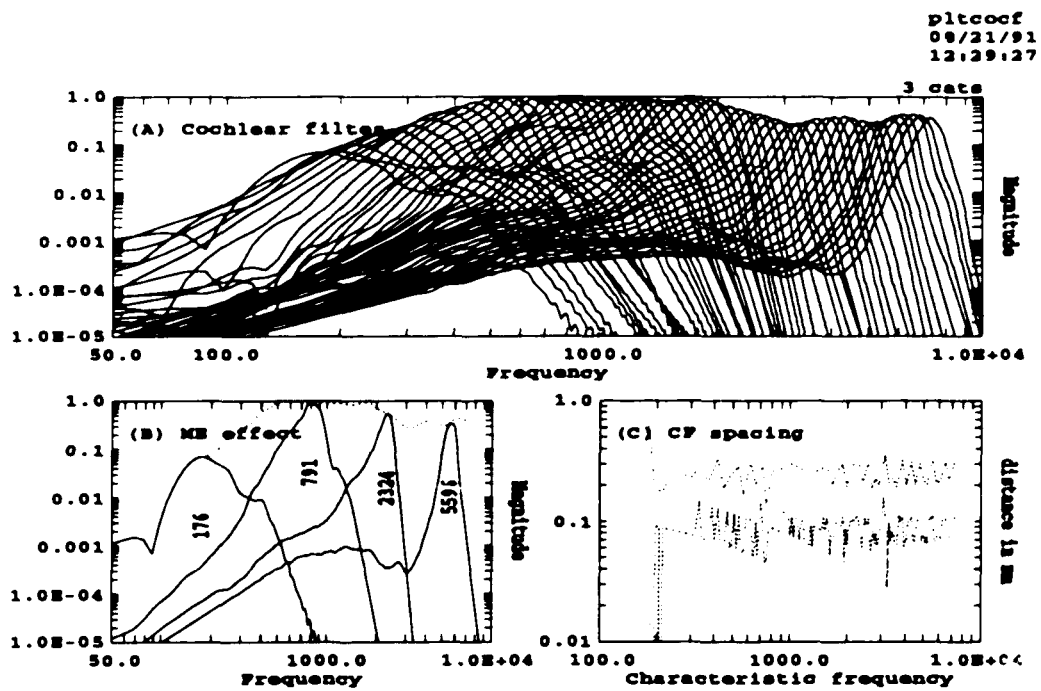


Figure 4.2: Original Cochlear Filters Used in EIH Model. In panel (A), 56 of 166 cochlear filters are shown (every third one). In panel (B), four cochlear filters are shown to demonstrate (a) different shapes of filters, and (b) the middle ear effect on the low frequency tails of filters with high characteristic frequencies (CF). The dashed line outlines the overall envelope of the cochlear filters at the center frequencies. In panel (C) The distance in mm between the CF's of every adjacent pair of cochlear filters is shown. The top dashed line shows the distances for the set of 56 filters given in (A) and the lower dashed line shows the distances for the extended set of 166 filters.

Hz and that of the highest filter is 7246 Hz. The filters are equally spaced on a logarithmic scale. The distance in mm between filter pairs is shown in panel (C) of the diagram. The upper dashed line shows the spacing for the original set of 56 filters. The distance between adjacent filters is mostly between 0.2 and 0.3 mm. The lower dashed line shows the spacing for the extended set of 166 filters. In this case the spacing between adjacent filters is between 0.05 and 0.15 mm, giving 7 to 20 filters per mm.

The characteristics of the middle ear transfer function are superimposed on the cochlear filters. The influence is apparent in two areas: (1) the shape of the overall envelope of the filter set, and (2) the individual filter shape in the tail region for the filters with CF's above approximately 2000 Hz. The overall envelope refers to a curve that connects the maximum-gain point of each filter (this maximum-gain point also identifies the characteristic frequency of the filter). The dashed line in panel (B) in Fig. 4.2 shows the overall envelope of the cochlear filter set shown in panel (A). The shape of the overall envelope shows that the transmission of signals in the 1-kHz region is most efficient, and that signals are attenuated below 500 Hz. There is also attenuation above 30 kHz (not shown in the diagram). The effect of the middle ear (ME) on the individual filters are easily observed among the filters with CF's above 1 kHz. These filters have shallow low-frequency tails that reflect the shape of the overall envelope of the filter set, hence they reflect the shape of the middle-ear pressure transfer function. The amount by which the filter shape is affected by the ME transfer function depends on the CF of the filter. The higher the CF, the more the tail reflects the ME effect. Three examples are given in panel (B). The low-frequency tail of the filter with CF at 791 Hz does not show the ME effect. For the filter with CF at 2324 Hz, its low-frequency tail from 300 Hz to 1 kHz parallels the overall envelope in

the same frequency region, indicating the effect of the ME. Likewise, the low-frequency tail of the filter with CF at 5596 Hz from 300 Hz to 2 kHz parallels the overall envelope in the corresponding frequency region, indicating the ME effect. The human ME pressure transfer function obtained from cadavers (Goode, 1987; Rosowski, 1991) seems to have less gain by 10 to 15 dB and appears flatter, i.e. it has shallower roll-offs on both low and high frequency sides (Puria, 1992.)

A shortcoming of the model is that the cochlear filters assumed for the model are linear in that the filter shapes are independent of the intensity levels. Rhode (1971, 1978) and others, however, have shown that the frequency response of the basilar membrane is level dependent. Cochlear models incorporating nonlinearities can be extremely complex and as a first step towards investigating physiologically based loudness models it is of interest to determine how well a simpler model, assuming linear tuning curves can work.

Half-wave Rectification The input-output function of IHC transduction is approximated by half-wave rectification in the model. This assumption is based on the observation that IHC's respond mainly to the depolarizing direction rather than to the hyperpolarizing direction (see Page 16). It should be noted that the input-output function of the IHC in the depolarization direction shows a nonlinear relation between the intracellular voltage change and the pressure change (Cody and Russell, 1987). The voltage changes rapidly initially with the pressure, and then saturates, resulting in a dynamic range of 30 dB. Therefore, the shortcoming of using half-wave rectification is that the rectifier does not accurately represent the saturating nonlinear property of IHC transduction. As noted above, a model incorporating all of the observed nonlinearities in the peripheral auditory system would be exceedingly complex and it is of interest to consider how well a model consisting of only

the simplest nonlinearity (half-wave rectification) can work.

Level detectors Five level detectors in each channel are used to simulate various thresholds of synapses at the base of an inner hair cell (see page 17.) Each level detector examines whether the filtered and half-wave rectified stimulus exceeds the threshold of the level detector. When the threshold of the level detector is exceeded, the level detector is activated. The time of activation is recorded for the next simulation step in the model. The thresholds of the five level detectors in each channel are spaced logarithmically over a pressure range that varies with the CF of the channel, as shown by the five symbols in Fig. 4.3. The channel with the lowest CF has a range of 60 dB and the channel with the highest CF has a range of 20 dB. The range decreases with increasing CF. The number of level detectors activated depends on the amplitude of the band-passed stimulus. Since the low-CF channels have larger ranges than the high-CF channels, the five thresholds in the low-CF channels are spaced further apart than in the high-CF channels. For the same reason, the rate of activation with increasing amplitude of the stimulus is slower in the low-CF channel than in the high-CF. In other words, the rate of activation increases with the CF of the channel. However, since there are only five level detectors, above the highest threshold of the level detector in each channel, increasing amplitude of a stimulus will not activate more detectors.

Physiologically, *each level detector represents a group of auditory nerve fibers of similar CF's and of similar sensitivity.* When the threshold of the level detector is exceeded, this group of fibers is *excited* or *activated*. Thus the thresholds of the auditory fibers are simulated by the model. As the amplitude of the stimulus increases, more fiber groups are excited. However, according to the design of the model, the fibers of low-CF activate

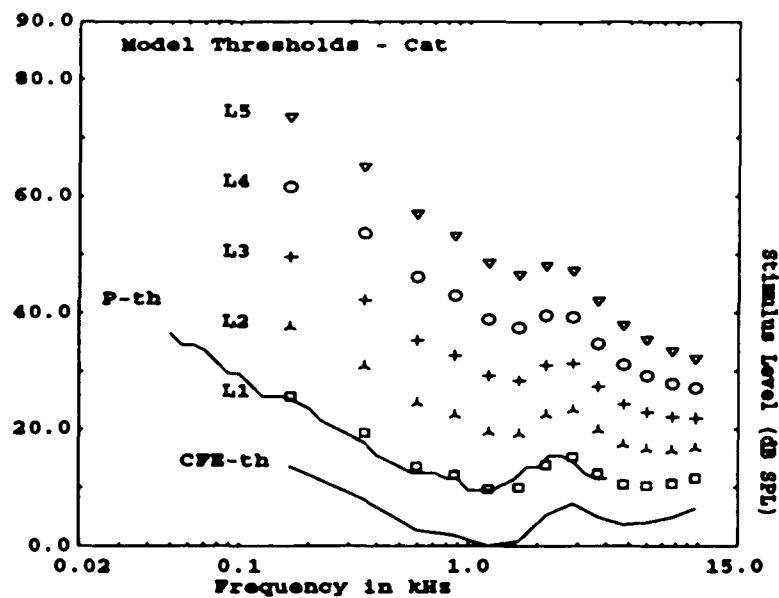


Figure 4.3: Sensitivity Thresholds of the Level Detectors in the EIH Model. Five threshold levels (L1 through L5) are shown as a function of frequency. Also shown is the predicted detection threshold (P-th) of the model and the envelope of the cochlear filter set (CFE-th), which is also the sensitivity thresholds of filters at CF. The predicted detection threshold is strongly affected by the envelope shape of the cochlear filters.

at higher threshold levels and at a slower rate of excitation with increasing stimulus level than the fibers of high-CF.

The rate-level function of an auditory nerve fiber is simulated by the EIH model in a very simplified manner. Since each level detector has only two states, i.e. the level detector is either activated or not activated, the group of fibers simulated by the level detectors has only two states, excited or not excited. The model ignores the noise in spontaneous firing at the threshold and the details of the rate function with intensity.

The phase-locking property of the auditory nerves is also taken into account by the level detectors. Given a pure tone as the input signal, when the half-wave rectified signal exceeds the threshold of a level detector, only the positive half of a cycle results in an excitation. As a consequence, excitation generated at the level detector occurs only once during a cycle and hence is phase-locked to the tonal stimulus.

Interval Histograms As mentioned earlier, the phase-locking property of the firings of the auditory nerves has been suggested to be another way of conveying the frequency information of a signal. In order to simulate the phase-locking property, the interval histogram was used to record the intervals between two adjacent activations of each level detector. However, it is frequencies of the intervals, the reciprocals of the interval period, that is of interest. The histogram is divided into 256 equal-Hz bins up to a maximum of 4000 Hz. Hence, the interval histogram, in essence, is a frequency histogram. One interval histogram is obtained for each level detector using the times at which the level detector is activated. The reciprocal of the interval gives the frequency, and then the bin of the histogram corresponding to the frequency is incremented. Each bin is 15.625 Hz wide. The first bin of the interval histogram corresponds to intervals greater than or equal to

64 msec ($1/15.625 \text{ Hz} = 64 \text{ msec}$). The second bin of the interval histogram corresponds to the intervals between 64 msec and 32 msec ($1/(2 * 15.625 \text{ Hz}) = 32 \text{ msec}$), and so on. The 256th bin corresponds to the intervals between .251 msec ($1/(255 * 15.625 \text{ Hz}) = .251 \text{ msec}$) and .25 msec ($1/(256 * 15.625 \text{ Hz}) = .25 \text{ msec}$ where $256 * 15.625 \text{ Hz} = 4000 \text{ Hz}$). For example, a single period of 1 msec causes the bin of 1000 Hz (frequency = $1/0.001 \text{ msec}$) to increase by 1. Since the activations occur on positive excursions of the signal (due to the half-wave rectification), the recorded activation times are phase-locked to positive half-cycles of a tonal signal. The frequency of a tonal signal is shown by a large bar in the interval histogram at a bin corresponding to the reciprocal of the period of the tone. One histogram is produced every 1.6 msec from the start of the stimulus; after every 1.6 msec, the reference time (time=0) for the histogram is reset. Each interval histogram in a sequence of histograms for a level detector is defined to be a frame.

Temporal integration is simulated by allowing a maximum of 20 adjacent intervals re the reference time for each interval histogram. A tonal signal activates a level detector at equal time intervals. Since a maximum of 20 intervals on each level detector are allowed, the 20 intervals are concentrated in one bin, the one corresponding to the frequency of the tone. A non-tonal signal activates a level detector at varying intervals, so the 20 intervals of the level detector will be distributed in various bins. In either case, the maximum overall value each interval histogram can have is 20. *This mechanism of processing the most recent 20 intervals on each level detector for each time frame is intended to simulate temporal integration.*

Additional rules for specifying integration time were introduced by Ghitza (1986, 1987, and 1988). One additional rule imposed an integration time of 40 msec for frequencies

below 500 Hz (for frequencies above 500 Hz, the maximum of 20 intervals per interval histogram (per level detector) determines the maximum integration time, which is equal to $20 * 1000/f$ milliseconds, where f is the frequency). A third rule limited the integration time further in a variable manner. From the first period to the 20th period, for each period the total elapsed time up to the current period is compared to 20 times the current period. If the elapsed time is greater than 20 times the current period, the remaining periods (the next period to the 20th period) are discarded.

The first two rules affect both pure tones and complex stimuli, the third rule affects complex stimuli only. Concerning the first two rules, if the input stimulus is a 1-kHz tone, then the integration time associated with each level detector, independent of cochlear filter, activated by the tone is 20 msec, i.e. 20 times 1-msec intervals. If the input stimulus is a 200-Hz tone, then the integration time associated with each level detector activated by the tone, is 100 msec, i.e. 20 times 5-msec intervals. Since 100 msec is greater than 40 msec, the integration time for every level detector is limited to 40 msec, which corresponds to 20 periods of a 500-Hz tone. Thus, the first two rules imply that the temporal integration time for each level detector is 40 msec for tonal stimuli below 500 Hz, and is 20 times the period of the tone for tonal stimuli above 500 Hz, e.g. 40 msec for 500-Hz tone, 20 msec for 1-kHz tone, and 10 msec for 2-kHz tone.

The third rule limiting integration time is more complex. It causes the integration time to vary from level detector to level detector, from channel to channel and from frame to frame. Following is an example of how the rule works for one level detector. Let the activation times of the last 20 periods relative to the onset of the signal be t_i , where $i = 0, 1, \dots, 21$. At the k th activation, if $t_k > 20 * (t_k - t_{k-1})$, then only the last $k - 1$

time intervals, given $t_{k-1} < 40$ msec, are included in the interval histogram. For the case of a pure tone, all 20 time intervals have the same period, hence 20 periods are used for the interval histogram. For this case, the first and third rule are identical. Consider now a complex stimulus, in which four 4-msec intervals are followed by a 0.75-msec interval. After the four 4-msec intervals, the time elapsed is 16 msec which is greater than 20 times the 0.75 msec interval. Thus, according to the third rule, the integration time is limited to 16 msec. Using the nomenclature introduced above, at the 5th activation $t_5 = 16.75$ msec and $t_5 > 20(t_5 - T_4)$ and only intervals t_1 through t_4 are included in the histogram.

A consequence of the above rules is that for complex stimuli, such as speech, the number of the intervals used for the histogram depends on the shape of the filter and the frequency components of the signal in a given channel. Further, the effect of the third rule, as illustrated by the above example of several large time intervals followed by very small intervals, is that smaller time intervals are more likely to be discarded; i.e., the high-frequency components in complex stimuli are more likely to be omitted from the interval histogram than the low-frequency components.

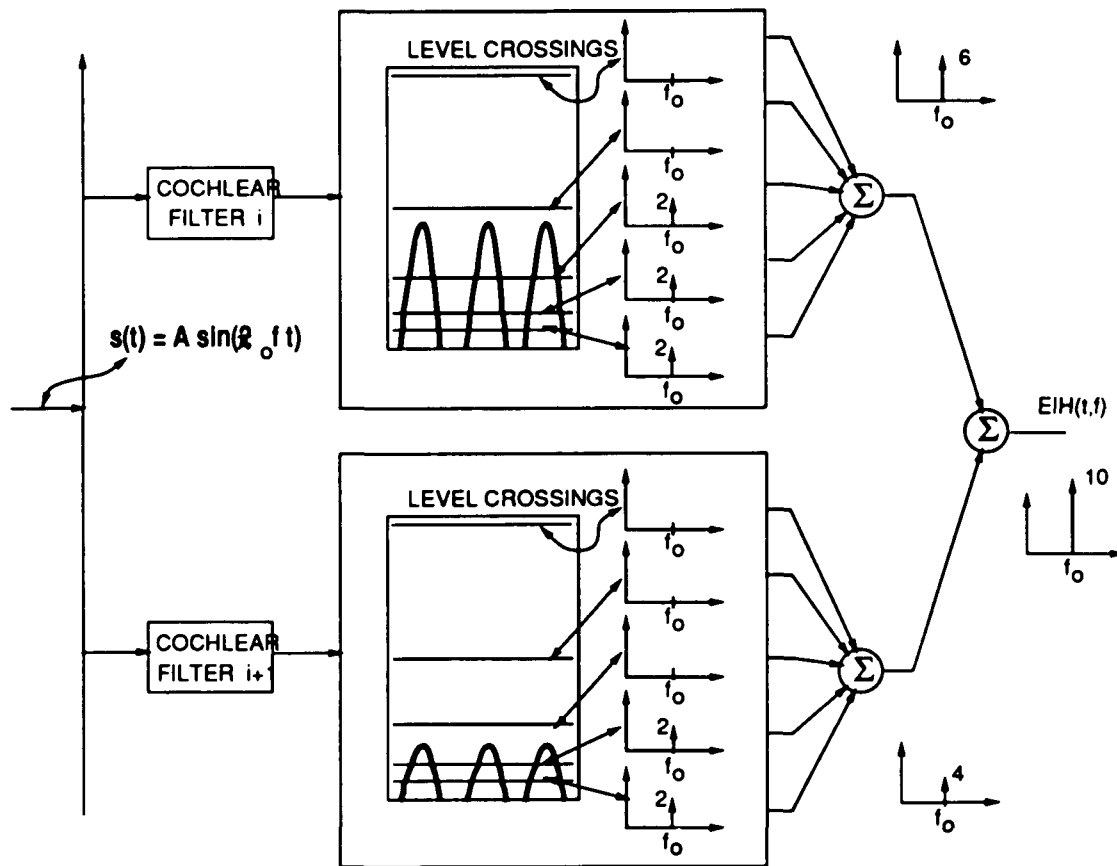
A rectangular weighting function is assumed for the temporal integration function; i.e. the weighting is constant for the duration of the time window and zero beyond. The temporal integration component of the model is of course heuristic, but the first rule is roughly consistent with psychophysical observations of temporal loudness summation observed by some workers (Sivian and White, 1933; Plomp and Bouman, 1959; Small, 1962; Sheeley and Bilger, 1964; Watson and Gengel, 1968.)

Ensemble Interval Histogram The *ensemble interval histogram* is the cumulative interval histogram summed across all filters and across all level detectors. The purpose of

this summation is to simulate the temporal synchrony that has been observed in auditory nerve fibers. If there is any correlation of periods among all fibers, the correlated period will stand out in the ensemble interval histogram. Thus the frequency content of the stimulus over the integration interval will be conveyed by the ensemble interval histogram. As mentioned in the previous paragraph, the histogram is expressed in period frequency. The maximum frequency of the interval histogram is limited to 4 kHz because cat neural data show that phase-locking is observed only up to 4 kHz.

The ensemble interval histogram is generated every 1.6 msec. Thus the output of the EIH model for a stimulus is a time sequence of ensemble interval histograms. The number of 1.6 msec is chosen because it is small enough so that when the stimulus being processed is complex, e.g. speech, the fine time structure of the stimulus will not be lost. The EIH model is also referred to as a *non-place* model because once the interval histograms are summed across all channels the place information along the simulated BM is lost.

Diagram of the EIH An illustrative diagram, Fig. 4.4, shows how the EIH model processes a pure tone, $s(t)$. At the first stage, the time wave form is filtered by the cochlear filters. The filter with a characteristic frequency nearest to the frequency of the tone passes it with nearly full amplitude, while other filters pass the tone at reduced amplitude. (See for reference the cochlear filter set in Fig. 4.2.) The output of each filter is then half-wave rectified. The number of level detectors activated depends on the amplitude of the half-wave rectified signal. In the example given, three of the five level detectors are activated and each detector is activated three times giving 2 time-intervals. Since the input stimulus is a pure tone, all two intervals have the same time period, and the inverse of the period gives the frequency of the tone, f_0 . Summing up the interval histograms for the five level

Figure 4.4: An Illustrative Diagram of EIH Model and Its Processing of a Pure Tone, $s(t)$.

detectors in the channel then gives a count of 6 for this channel. Finally, the interval histograms for all the channels are summed to give an ensemble interval histogram.

Excitation Patterns Excitation patterns at various stages of the model are useful and important in later discussions, therefore they will be described next. The excitation patterns at two stages are described in the following: the excitation amplitude as a function of place along the BM is referred to as *the BM excitation pattern*, and the excitation of the auditory nerves as a function of the CF or of the place on the BM corresponding to the CF is referred to as *the neural excitation pattern*. In the EIH model, the BM excitation pattern is expressed as the amplitude of the signal passed by each filter as a function of place along the basilar membrane. The neural excitation pattern of the model is expressed as the total number of times the five level detection thresholds are activated as a function of channel CF's. Since each level represents a group of neurons of similar CF and of similar threshold level, the excitation pattern shows the total number of neurons excited, averaged over the integration time, as a function of position along the basilar membrane. In the EIH model, the *spread of excitation* accompanying the increasing amplitude of the input signal at the basilar membrane is expressed by the increasing number of channels being activated.

An ensemble interval histogram can also be considered to be a simulated neural excitation pattern in the EIH model, since the histogram carries the same information as the neural excitation pattern defined earlier. The difference of the histogram from the neural excitation patterns is that the period information of the neural activity in the latter is extracted from and collapsed over all neurons over the integration time, and is expressed as a function of the inverse period, i.e. frequency in Hz.

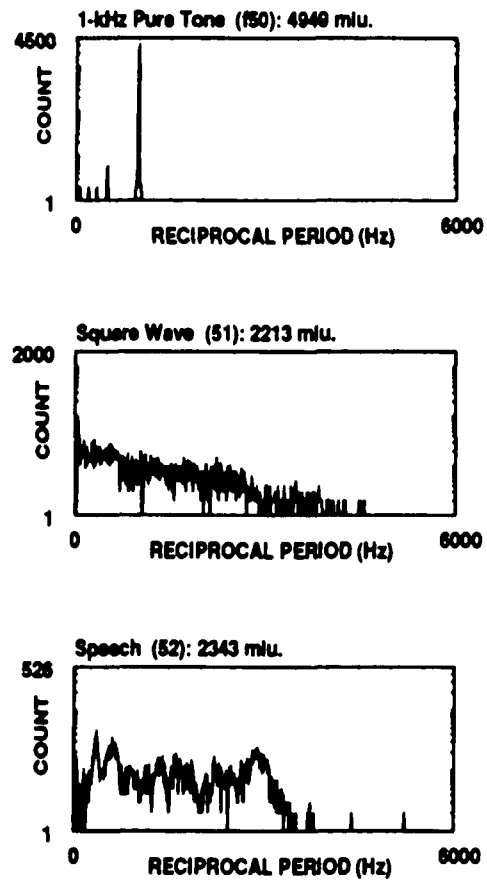


Figure 4.5: [E

Examples of EIH output Examples of EIH output for three different input stimuli: 1-kHz tone, square wave, and speech. In the title of each panel, the stimulus type (e.g. 1-kHz Pure Tone) is followed by the frame number (e.g. f50), which is followed by the total loudness units of the stimulus by the EIH-based loudness model (e.g. 4949 mlu), where 'mlu' will be defined in the next section.

Examples of ensemble interval histograms are given in Fig. 4.5 for three different inputs: 1-kHz tone, square wave, and speech. The ordinate of a histogram represents the number of the occurrences of various time intervals and the abscissa represents the inverse of time interval, $1/\Delta t$, defining a spectrum. Another way of looking at the ensemble interval histogram is that a point (x, y) on the histogram means that an interval period of $1/x$ second occurred y times. When the input is a pure tone, the firing patterns will be periodic with the same period as the input pure tone, and the ensemble interval histogram will have a component at the interval bin that corresponds to the period of the tone.

4.2 The Loudness Module in the EIH-LM Model

In the EIH based loudness model (EIH-LM), the *instantaneous loudness* is defined to be proportional to the total number of intervals in one ensemble interval histogram frame, where the total number of intervals is summed across all level detectors and all channels over the integration time. In terms of the physiological correlates used in the EIH-LM model, the instantaneous loudness is proportional to the total number of groups of neurons excited and the number of excitations over the integration time period. Note that the instantaneous loudness can also be viewed as a short-term loudness since it is summed over the temporal integration time in one frame, which is generated every 1.6 msec, a short

term.

The *overall loudness* of a signal over time is defined to be the mean instantaneous loudness averaged over the number of ensemble interval histogram frames for the duration of the signal. The loudness computed for a signal is also referred to as *predicted loudness* in the following discussions. The unit of the instantaneous loudness or overall loudness is defined to be the *model loudness unit* (mlu).

Loudness in sones can be predicted by the model using the same strategy used in defining the sone scale. The predicted loudness of a 1-kHz tone at 40 dB SPL is defined to be 1 sone. The sone value of any stimulus can be obtained by dividing the model loudness units (the predicted loudness) of that stimulus by the model loudness units calculated for a 1-kHz tone.

The prediction of loudness levels by the model is obtained in a manner analogous to that used in measuring loudness levels in phons. When two stimuli (*A* and *B*) are matched for equal loudness, the model first predicts the loudness in mlu for the fixed stimulus. The loudness in mlu for the adjusted stimulus is then computed for several stimulus levels. The level of this stimulus is then adjusted until its loudness in mlu is equal to that of the reference stimulus.

Detection Thresholds

The detection threshold for a stimulus predicted by the EIH-LM model is defined to be the lowest sound pressure level at which the predicted loudness is greater than or equal to 20 mlu, the number defined in Section 4.1 to be the maximum temporal integration time, during which time period the temporal loudness sums. This number refers to the

maximum number of excitations a level detector can contribute to the total loudness in an EIH frame. For 1-kHz pure tone at the threshold level, only the cochlear filter tuned to 1 kHz will respond to the tone. In this filter, only the level detector with the lowest sensitivity threshold will be activated. However, only when the first 20 excitations is reached, is the loudness fully established and will the predicted instantaneous loudness of the stimulus remain constant. The same is observed in measuring detection threshold for human, i.e., the detectability of a stimulus increases with the duration of the stimulus and the ear requires time for a signal of a fixed sound pressure level to establish its loudness.

4.3 Twelve Versions of the EIH-LM Model

In the process of evaluating the performance of the EIH-LM model in loudness prediction, several parameters of the model were modified. In order to distinguish one version of EIH-LM from another, the version number is attached to the end of abbreviation; e.g., EIH-LM1 for version 1. The modifications of each version are described next.

EIH-LM1 The EIH-LM model described in Sections 4.1 and 4.2 is defined as EIH-LM1.

EIH-LM2 The EIH-LM2 differs from the EIH-LM1 in that the 40-msec temporal integration time limit is removed so that the last 20 intervals are used for all pure tones to construct the interval histograms.

EIH-LM3 EIH-LM3 differs from EIH-LM2 in the cochlear filter set. A new second set of 71 cochlear filters was derived from a set of cat tuning curves with CF ranging from 163 Hz to 15.6 kHz². The new filter set has filters with higher CF's, as shown in Fig. 4.6.

²This set of cat tuning curves was provided by Delgutte from Eaton Peabody Laboratory.

Cochlear filters of high CF are activated by lower frequency components of a sound at and

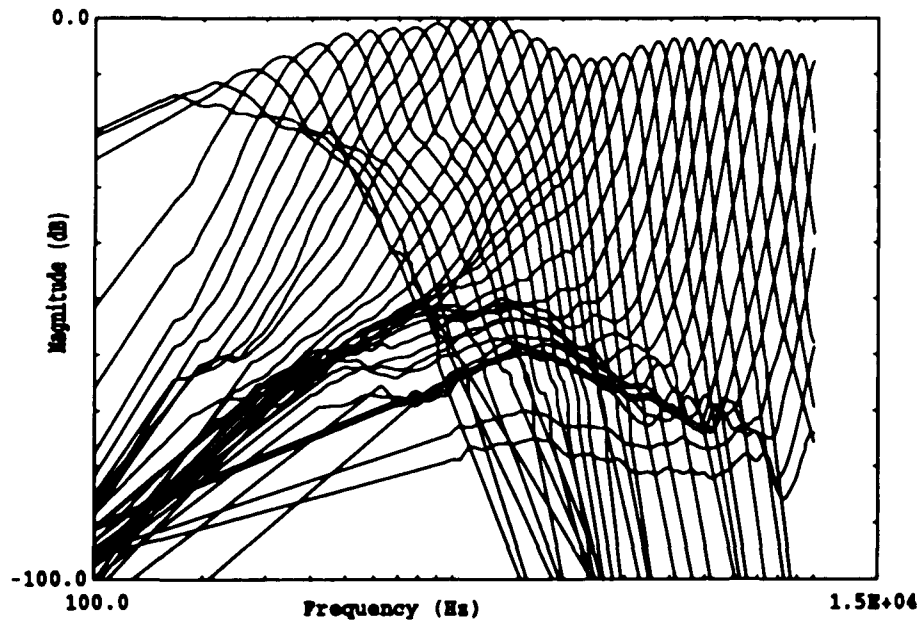


Figure 4.6: New Set of Cochlear Filters Used in the EIH-LM Models. The second set of 71 cochlear filters derived from cat neural tuning curves is used in the EIH-LM models from version 3 to version 12. The characteristic frequencies of the filters range from 163 Hz to about 12 kHz. The highest characteristic frequency in the first cochlear filter set shown in Fig. 4.2 is 7246 Hz while that for the second filter set is extended to 12 kHz. However, due to signal processing limitations, portions of the filters with CF above 10 kHz are truncated at 10 kHz. Only every other filter is plotted to make viewing easier.

above 70 dB SPL due to the shallow low-frequency tail of those filters. Thus, the presence of cochlear filters with high CF should have an effect on the calculated loudness above 70 dB SPL. The filter set was thus extended in order to examine the effect of the inclusion of high-CF filters on predicted loudness at lower frequencies.

EIH-LM4 EIH-LM4 differs from EIH-LM3 in the definition for the overall loudness. In the previous versions (1 to 3), the overall loudness of a stimulus is the mean instantaneous

loudness, the loudness computed for one histogram, averaged across all ensemble interval histograms. In EIH-LM4, the overall loudness of a stimulus is the loudness value such that 90% of all instantaneous loudness values are below it, or, 10% of the times, the instantaneous loudness values are above it. To formulate the two computations of the overall loudness values, let L_i represent the instantaneous loudness calculated for the i th histogram, where $i = 1, \dots, n$, and n = the total number of histograms for the stimulus. Let OL stand for the overall loudness value, then the OL for EIH-LM3 equals $OL = 1/n \sum_{i=1}^n L_i$. To compute OL for EIH-LM4, the first step is to find the m that satisfies the inequality: $(\sum_{j=1}^{m-1} L_j)/(\sum_{i=1}^n L_i) < 0.90$, or alternatively $(\sum_{j=m}^n L_j)/(\sum_{i=1}^n L_i) \geq 0.10$, where $m \leq n$. Then L_m is the overall loudness of the stimulus, based on the percentile method.

The above definition of the overall loudness should not affect the pure tones, because the instantaneous loudness for a pure tone is constant from frame to frame. Therefore the two methods of predicting the overall loudness should yield the same results. However, in the case when the dynamic range of the stimulus varies widely and rapidly, the instantaneous loudness also varies likewise, and hence the two methods will yield very different results. Examples will be given in Chapter 7.

EIH-LM5 EIH-LM5 differs from EIH-LM4 in the total range of the threshold levels of the five level detectors. Instead of having variable threshold range with the channels, the threshold range is fixed at 60 dB for all channels. The five threshold levels for the five level detectors are distributed at equal logarithmic steps (12 dB). This modification indicates that regardless of the CF's of the channels, the neurons in each channel have similar sensitivity property. In other words, there are same number of neurons for each

sensitivity level per channel, and the sensitivity levels of the neuron are the same regardless of the channel.

EIH-LM6 EIH-LM6 differs from EIH-LM5 in the total sensitivity threshold range and the number of level detectors per channel. In EIH-LM6 the total range is expanded from 60 dB to 80 dB and the number of level detectors is increased from 5 to 8. This modification alters the model in the direction of expanded rate-level functions instead of using all-or-none rate functions.

EIH-LM7 EIH-LM7 differs from EIH-LM6 in the third temporal integration time. In EIH-LM7, the third temporal integration rule is removed, so that the original temporal integration rule is the only rule remaining. Regardless of tonal or complex stimulus, 20 intervals are used for every level detector in every channel to determine the temporal integration time.

EIH-LM8 In EIH-LM8, the number of level detectors was doubled from EIH-LM7 with the range of threshold levels remaining the same. The threshold levels of the level detectors were spaced equally logarithmically, i.e. equal number of dB between adjacent detectors.

EIH-LM9 and EIH-LM10 The main difference of EIH-LM9 and EIH-LM10 from the previous versions was in the way the thresholds of the level detectors were spaced. Instead of equally spaced in logarithmic scale, i.e. dB, they were equally spaced linearly, i.e. in pressure. EIH-LM10 had a higher density of level detectors than EIH-LM9.

EIH-LM11 and EIH-LM12 Both EIH-LM11 and EIH-LM12 had 13 level detectors. In EIH-LM11, the sensitivity thresholds of level detectors were arranged in decreasing thresholds spacing in dB starting at 12 dB, i.e. 12 dB between the thresholds of the first two level detectors, 11 dB between the thresholds of next two level detectors, and so on until 1 dB is used between the thresholds of the highest two level detectors. In EIH-LM12, the arrangement of the thresholds of the level detectors is just the opposite. The difference between the lowest two level detectors is 1 dB, the difference between the next two lowest level detectors is 2 dB, and so on until the difference is 12 dB between two thresholds. The configuration in EIH-LM11 caused the thresholds of level detectors to cluster at high SPL, and that in EIH-LM12 caused the thresholds to cluster at low SPL.

Chapter 5

EIH-LM Prediction I

As demonstrated by the loudness functions, the loudness level functions, and the equal-loudness contours in Chapter 3, the loudness of a pure tone is a nonlinear function of its sound pressure and frequency. In the first part of this chapter, the first three versions of the EIH-based loudness model (EIH-LM1, EIH-LM2, and EIH-LM3) were evaluated by comparing the predicted loudness for pure tones against observed loudness data for humans. The evaluation shows that EIH-LM3 gives better predictions than EIH-LM1 and EIH-LM2. In the second part of the chapter, the EIH-LM3 model was evaluated further by predicting critical band data using loudness summation of a four-tone complex as reported by Scharf (1959). In this experiment, Scharf examined the critical bandwidth of a constant-energy four-tone complex of variable bandwidth for four near-threshold levels at four center frequencies.

The three versions of the loudness model have been described in detail in Section 4.3. For convenience, the differences among three versions are briefly summarized. EIH-LM2 is the same as EIH-LM1 with 40-msec integration rule removed. EIH-LM3 is the same as EIH-LM2 except EIH-LM3 uses a new set of cochlear filters with the highest characteristic

frequency extended to 12.5 kHz, whereas EIH-LM1 and EIH-LM2 used cochlear filters with a range up to 7.2 kHz.

5.1 Predictions for Pure Tones

5.1.1 Loudness Functions

In this section, pure-tone loudness functions predicted by EIH-LM1, EIH-LM2, and EIH-LM3 are compared to those measured in humans¹. The observed loudness functions reported by Scharf (Table I and Figures 1 & 5, 1978) and by Hellman (1976) are used for these comparisons.

Comparisons for five tone frequencies are shown in Fig. 5.1. In panels (a), (b), and (c), the predicted loudness functions for 250, 500, and 1000 Hz are compared to the observed loudness functions at those frequencies (Scharf, 1978). In panels (d) and (e), the predicted loudness functions for 2 and 3 kHz are compared to the observed loudness function at 3 kHz (Hellman, 1976). The observed 3-kHz loudness function is used to compare with the predicted 2-kHz loudness function because the observed loudness function for a 2-kHz tone is not available and the observed loudness function for 3-kHz is similar to that for 2-kHz tone, as can be deduced from the iso-loudness contours in Fig. 3.2. In each panel, the abscissa represents the sound pressure level of the tone. The ordinate represents the loudness in sones. The open circle in (c), at coordinate (40,1), indicates the reference condition used in deriving the sone scale for both the observed and predicted loudness functions. By definition, both observed and predicted loudness functions must pass through this point.

¹Since this study is concerned only with human loudness data, the term *observed loudness* rather than human loudness data will be used from this point on.

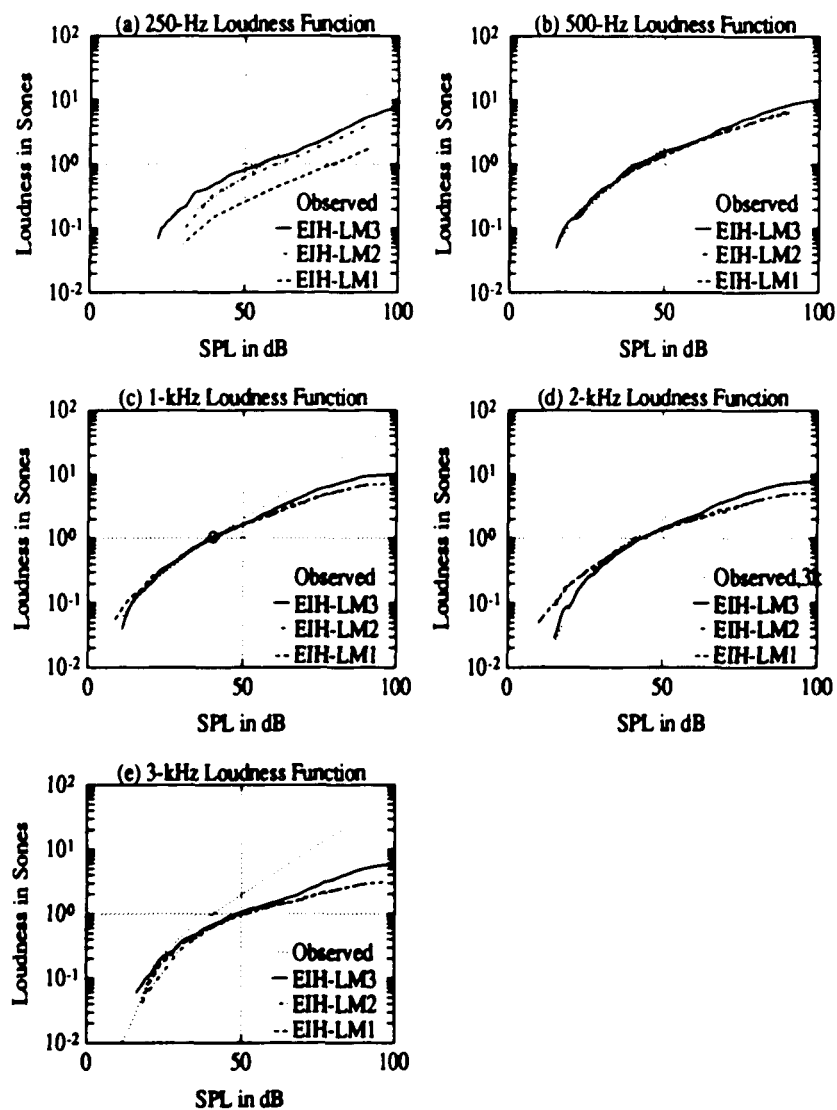


Figure 5.1: Comparison of the Observed and Predicted Pure Tone Loudness Functions. Pure tone loudness functions predicted by EIH-LM1, EIH-LM2, and EIH-LM3 are compared to those measured in humans. Loudness functions calculated for pure tones at 250, 500, and 1000 Hz are compared to the observed loudness functions at the corresponding frequencies. Loudness functions calculated for 2 and 3 kHz are compared to the observed loudness function at 1 kHz. In each panel, the abscissa is the sound pressure level in dB. The ordinate represents the loudness in sones. The loudness functions predicted by EIH-LM1 and EIH-LM2 are almost identical for frequencies 500 Hz and higher.

Comparison between the loudness functions predicted by EIH-LM1 and those by EIH-LM2 show that the predicted loudness functions are almost the same except at 250 Hz. As shown in panel (a), the EIH-LM2 loudness function lies closer to the observed loudness function than the EIH-LM1 loudness function. At low dB SPL, the EIH-LM2 loudness function matches the observed loudness function rather well, except at 2 kHz. (However, one must bear in mind that the observed loudness function for comparison in (d) is 3 kHz. If 1 kHz were used for the comparison, not shown, the statement would be true for 2 kHz also.) At high dB SPL, the slope of the EIH-LM2 loudness function is shallower than the observed loudness function. The loudness functions predicted by EIH-LM3 differ from those predicted by EIH-LM1 and EIH-LM2 in several ways. At high dB SPL, except for 250 Hz, the slope of the EIH-LM3 loudness function becomes steeper. At 80 dB SPL, the EIH-LM3 predicted loudness is about 1.4 to 1.6 times that of EIH-LM1 and EIH-LM2, resulting in increase of the dynamic ranges of the predicted loudness functions. At low SPL, slight variations can be observed for frequencies above 1 kHz. At 2 kHz, the EIH-LM3 loudness function matches very well with the observed loudness function. The EIH-LM3 loudness function for 250 Hz is shifted left without too much change in slope above 50 dB SPL but with rather larger variation below 50 SPL.

Comparing to observed loudness functions, the predicted loudness functions for frequencies above 500 Hz are relatively accurate below 50 dB SPL, including the rapid reduction in loudness toward absolute threshold. Above 50 dB SPL, the slope of the predicted loudness function is much shallower than that of the observed functions. By increasing the number of cochlear filters with high characteristic frequencies, as in EIH-LM3, the slopes of the predicted loudness functions improve slightly and the functions are closer to the observed

loudness functions. The improvement also increases with frequency, i.e. the improvement is greater at 3 kHz than at 500 Hz. At 250 Hz, it is difficult to determine whether the EIH-LM3 function is an improvement, since the function overestimates loudness at low SPL, while the problem at high SPL remains.

Additional improvements in predicting loudness functions are possible by varying other parameters of the model. In Chapter 7, further modifications of the EIH-LM model are evaluated for both complex sounds and tonal stimuli.

Fig. 5.2 compares the slopes of the observed loudness functions with those of the predicted loudness functions shown in Fig. 5.1. The two figures have the same format except the ordinates of Fig. 5.2 show the slopes of the loudness functions in $20 * \log(\text{some})$ per dB SPL. The slope shown also represents the exponent of the power function relating the loudness of a sound in sones to its sound pressure in dynes per square centimeter. A slope of 1 indicates that the loudness grows in linear proportional to the sound pressure. As the slope decreases from 1, the loudness grows more slowly with increasing sound pressure, and when the slope remains at a constant, the loudness function appears as a straight line on a log-log scale. When the slope approaches 0, the loudness approaches saturation and no longer increases with increasing sound pressure.

At low SPL, both observed and predicted loudness functions show a rapid decrease in slope with increasing sound pressure level. In this region, the slopes of the three model versions vary widely. Above about 40 dB SPL, the slopes of the observed loudness functions stabilize at a constant between 0.6 to 0.7. The slopes of the predicted functions also reach a quasi-stable condition from 50 to 65 or 80 dB SPL depending on the frequency. The slopes under these conditions range from 0.25 to 0.45 depending on the frequency and

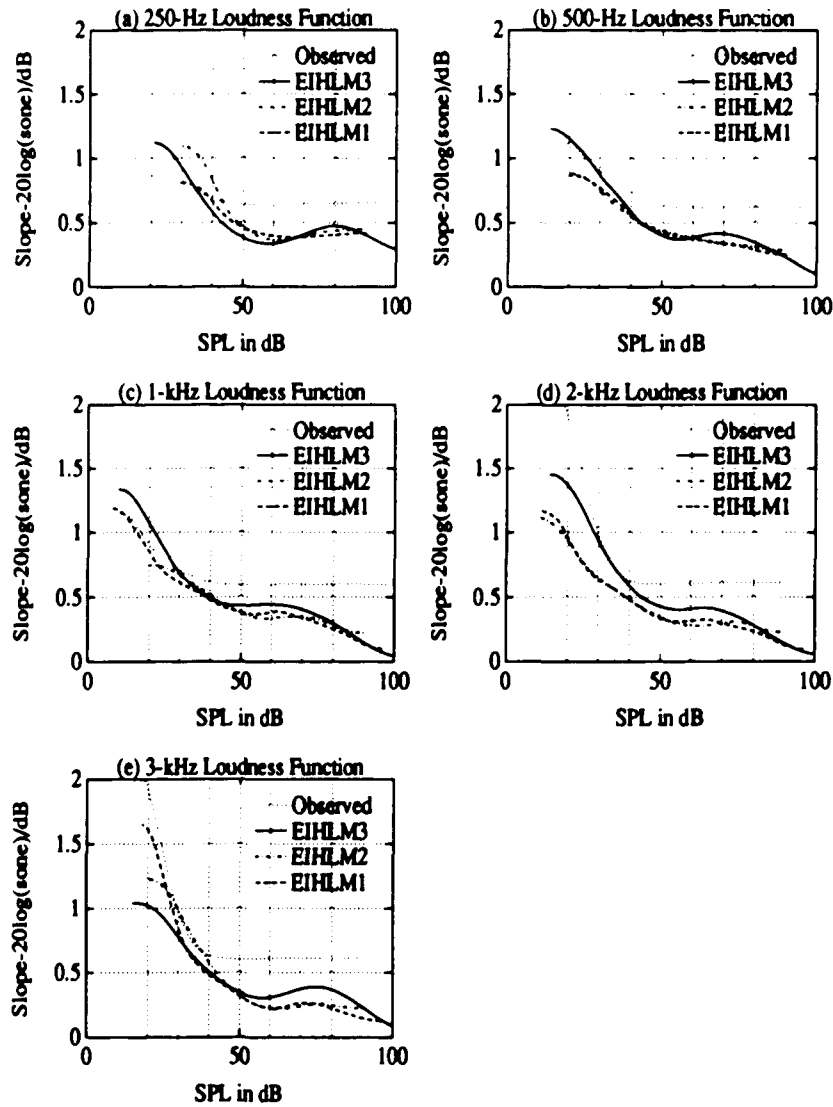


Figure 5.2: Slopes of the Pure Tone Loudness Functions as a Function of SPL are Shown. The same format as Fig. 5.1 is used except the ordinates represent the slopes of the loudness functions in $20 * \log(\text{sones})$ per dB SPL. The slope obtained in this way corresponds to the exponent of the power function relating loudness in sones to pressure in dynes per square centimeter.

the version of the model used in deriving the predictions. A comparison of the slopes for the three sets of predictions with the slopes of the observed loudness functions shows that EIH-LM3 model provides the best agreement, and that the relative quality of the EIH-LM3 predictions (relative to the EIH-LM1 and EIH-LM2 predictions) improves with frequency. At high SPL's, the slopes of the predicted loudness decrease toward 0, i.e. the predicted loudness functions begin to saturate at high SPL's, as shown in Fig. 5.1. The amount of saturation increases with frequency, and the saturation threshold decreases with increasing frequency.

Discussion

The improvement of the predicted loudness functions at 250 Hz due to the removal of the 40-msec integration rule seems to support the notion of an integration time that is related to frequency, i.e. an integration time determined by the number of cycles needed to establish the loudness. As a consequence, low frequency tones require a longer integration time and high frequency tones require a shorter integration time. In EIH-LM2 and EIH-LM3, an integration time corresponding to 20 periods was used for the pure tones, i.e., 100 msec for a 200-Hz tone, and 20 msec for a 1-kHz tone, and 10 msec for a 2-kHz tone. This frequency dependency of the temporal integration time is compatible with data obtained by Plomp and Bouman (1959), Watson and Gengel (1968), Sheeley and Bilger (1964), and others, although the actual integration times do not agree. The data obtained by different workers also vary because of the methods used and the wording of the instructions to the subjects.

The saturation of the predicted loudness functions at high sound pressure levels may be

a result of the restricted spread of neural excitation due to the limited set of the cochlear filters used in the model. Examination of the cochlear filter set in Fig. 4.6 shows that the filters excited by any tone can be divided into two groups, those with characteristic frequencies below the frequency of the tone and those above. For a fixed sound pressure level, the number of filters in the second group is greater for a low-frequency tone than for a high-frequency tone, because of the long low-frequency tails of the filters. For example, the filter with a characteristic frequency of 3 kHz has a low-frequency tail which can cause the filter to be excited by a tone of 500 Hz at 80 dB SPL. Thus the neural excitation pattern for a low-frequency tone can continue to grow until all the filters with higher CF are excited. While the neural excitation pattern for a high-frequency tone also grows with the intensity until all the filters with higher CF are excited, the number of filters available for neural excitation is lower than that for a low-frequency tone. As a consequence, the predicted loudness function saturates at a higher level for 250 Hz than for 500 Hz. Similarly, the predicted loudness function saturates at a higher level for 500 Hz than for 1 kHz, and so on. This also explains why the EIH-LM3 model, which includes cochlear filters above 7.2 kHz, provides better predictions of loudness functions at high SPL's.

Many workers (Hellman, 1978; Viemeister, 1988; and others) have argued that the spread of excitation alone is insufficient to explain the observed dynamic range of hearing. The effect on the predicted loudness functions of extending the filter set to include filters of characteristic frequency higher than 7.2 kHz shows that the spread of excitation along basilar membrane does contribute to the loudness growth. However, it alone is not sufficient to account for the constant loudness growth rate above 40 dB SPL. Above 80 dB SPL, the excitation pattern stops growing on the high-frequency side and grows very slowly on

the low-frequency side, as can be observed in Fig. 4.6. If one believes that the observed loudness functions represent loudness perception, then other parameters have to contribute to loudness growth. The neural firing rate of auditory fibers has been thought by many to be the parameter. More on this point will be discussed later.

In the first three versions of the EIH-LM model, the saturation problem is further compounded by the varying threshold range of the level-crossing detectors, larger in low-CF channels and smaller in high-CF channels, as shown in Fig. 4.3. The smaller threshold range in high-frequency channels causes the excitation pattern of a high-frequency tone at low SPL to grow at a slight higher rate with SPL than a low-frequency tone. This seems to correspond to the observed loudness functions. In the model, the increased growth rate at low level for a high-frequency tone occurs because the level detectors are excited at lower sound pressure levels and at smaller intervals between level detectors. However, this may not be the reason for rapid growth rate at low SPL of an observed loudness function for a high-frequency tone. From moderate to high SPL, the smaller threshold range causes excitation pattern for a high-frequency tone to grow at a much lower rate with SPL, i.e. the excitation pattern and hence the predicted loudness function has a shallower slope and saturates at a lower SPL. A comparison of all the predicted loudness functions from 250 Hz to 3 kHz (see Fig. 5.1 or Fig. 5.2) shows that the slope of the predicted loudness function above 50 dB SPL decreases with increasing frequency.

The overestimation of loudness by the EIH-LM3 model at low levels for 250 Hz may be an artifact of the extrapolation procedure that converted the neural tuning curves to the filters. At very low characteristic frequencies, the neural tuning curves did not have enough data points for extrapolating the low frequency slopes for the filters, hence the estimated

slopes may be too shallow and causes the predicted loudness function, e.g. 250 Hz, to grow faster at low SPL. An examination of the cochlear filter set in this region shows that the slopes of these filters are rather important in determining the loudness growth rate of a tone with very low frequency.

Another problem for loudness growth at lower frequencies, e.g. 250 and 500 Hz, is the lack of cochlear filters at low frequency, i.e. below a CF of 180 Hz. The lack of filters is due to the lack of cat neural tuning curves. The effect of including low-CF cochlear filters should be similar to the effect of including high-CF filters except the frequency region affected should be much smaller.

5.1.2 Loudness Level Functions

A loudness level function expresses the SPL of a reference 1-kHz tone judged to be equal in loudness to a test stimulus as a function of the SPL of the test stimulus. Fig. 5.3 shows the observed and predicted loudness level functions for pure tones at 250, 500, 2000, and 3000 Hz. The observed loudness level data are from Fletcher and Munson (1933). The predicted loudness level functions are for models EIH-LM1, EIH-LM2, and EIH-LM3. The dotted diagonal line is used only as a guide; any point on the line indicates that test and reference stimuli of equal SPL are also equal in loudness. The observed loudness levels are shown by the + symbols. In panel (d), two sets of observed loudness levels are presented, those for a 4 kHz test tone represented by + and those for an 8-kHz test tone represented by o. The observed loudness levels for 4-kHz are used because the data for 3 kHz are not available. The 8-kHz data are used because 3 kHz, being close to the upper frequency limit of the EIH-LM model, is like 8 kHz being close to the upper limit of the human audition. The parallel lies in that the total number of level detectors activated by a 3-kHz tone in

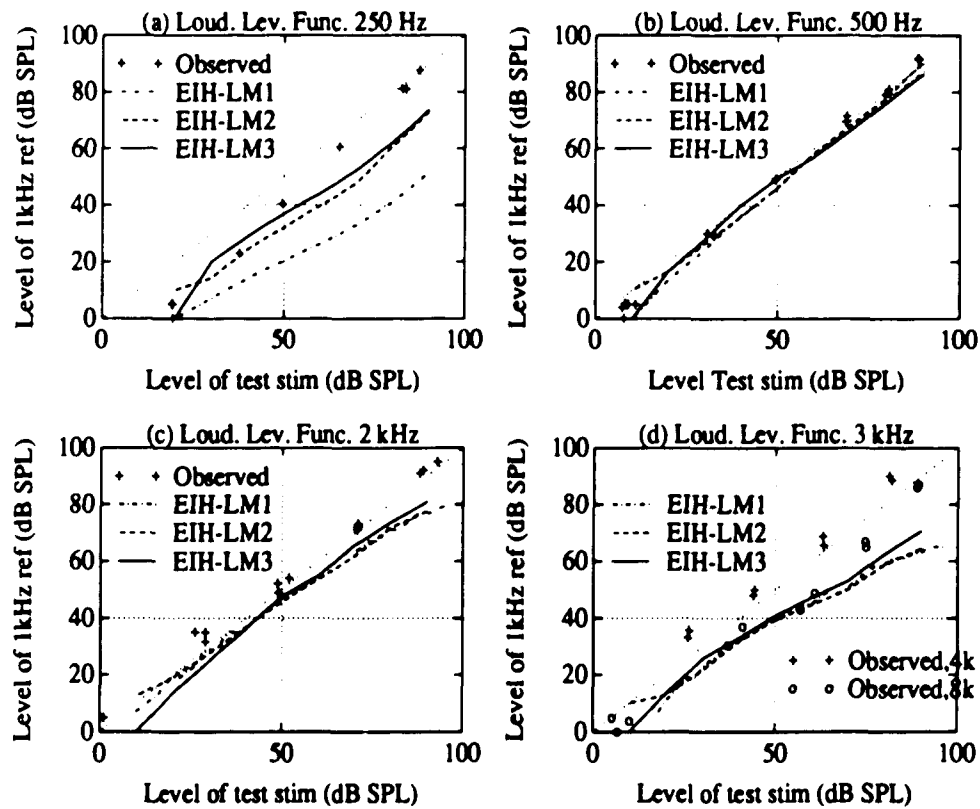


Figure 5.3: Comparison of the Observed and Predicted Loudness Level Functions for Pure Tones. Comparison of loudness level functions predicted by EIH-LM1, EIH-LM2, and EIH-LM3 to observed loudness level functions are shown for pure tones at frequencies, 250, 500, 2000, and 3000 Hz. The ordinate of each curve shows the SPL of 1-kHz tone judged to be equal in loudness to the test stimuli. The level of the test stimulus is shown on the abscissa.

the model and the neurons excited by an 8-kHz tone in human are limited by the upper bound of the frequency range. Therefore, the behavior of the observed loudness levels for an 8-kHz tone seem more appropriate than those for a 4-kHz tone for comparison to the predicted data at 3 kHz.

The slope of the loudness level function shows the rate of loudness growth of the test sound to the 1-kHz reference tone. If the slope of the loudness level function is the same as that of the dotted diagonal line, then the test sound has the same loudness growth rate as the reference tone. If the slope is shallower than that of the dotted diagonal line, then the loudness of the test sound grows at a slower rate than that of the reference tone. If the loudness level function is shifted to the right of the dotted diagonal line, then the threshold of the test sound is higher than that of the reference tone and, above threshold, the reference tone will sound louder than the test sound for the same SPL.

Fig. 5.3 shows that, unlike the loudness functions, the loudness level functions predicted by the three model versions have not changed the relations between the model loudness of 1-kHz and that of other frequencies significantly, except 250 Hz. In panels (b - d), the predicted loudness levels functions overlap each other for the most part.

The comparison of the observed and the EIH-LM3 predicted loudness level functions shows that: (1) At 250 Hz (panel (a)), the observed threshold of 250-Hz tone is higher than that of the reference tone. However, the slope of the observed loudness function is steeper than the diagonal dotted line, indicating that the loudness of 250-Hz tone grows faster than that of the reference tone. Above 85 dB SPL, the observed loudness level function merges with the diagonal dotted line, indicating that the 250-Hz tone is equally loud as the reference tone at the same sound pressure level and grows at the same rate as the reference

tone. The predicted loudness level function agrees roughly with the human function below 50 dB SPL. However, above 50 dB SPL, the model loudness continues to grow at about the same rate as below 50 dB SPL and does not increase to merge with the diagonal line.

(2) At 500 Hz (panel (b)), the predicted and observed loudness level functions coincide.

(3) At 2000 Hz (panel (c)), the predicted threshold of the 2-kHz tone is higher than that of 1-kHz tone, which is opposite to the observed data. Near threshold, the predicted loudness grows slightly faster for a 1-kHz reference tone than for 2-kHz test tone. Between 25 and 70 dB SPL, the predicted loudness levels are rather close to the observed loudness levels. Above about 70 dB SPL, the predicted loudness level function lies below the observed loudness level function and the slope of the predicted function is slightly shallower than that of the observed function. This indicates that the growth rate of the predicted loudness for 2 kHz is lower than the growth rate of the predicted loudness for the 1-kHz reference, unlike the case of the observed loudness level function which shows the growth rates of both 2-kHz tone and the 1-kHz reference are nearly the same.

(4) At 3 kHz (panel (d)), the predicted loudness level function matches the observed loudness level function at 8 kHz much better than that at 4 kHz. The predicted loudness levels and the observed loudness levels for 8-kHz tone coincide below 60 dB SPL.

Discussion

A loudness level function of a sound shows the loudness growth rate of a sound relative to that of a 1-kHz reference tone. The improvement of the loudness level function at 250 Hz due to the removal of the 40-msec temporal integration time limit reaffirms that the temporal integration time limit is a poor choice.

The extension of the cochlear set to include filters of higher CF, up to 12.5 kHz, does not seem to affect the loudness growth of the test tones relative to the 1-kHz reference tone, except near threshold for 2-kHz tones and at very high SPL for 2- and 3-kHz tones. At very high SPL, the loudness level functions predicted by EIH-LM3 do not become saturated as soon as those predicted by EIH-LM1 and by EIH-LM2, indicating that increased high-CF filters reduce the saturation problems for tones of higher frequencies. In this case, the improvement for 3 kHz is larger than 2 kHz and the improvement for 2 kHz is more than 1 kHz.

The differences between predicted loudness levels and observed loudness levels at 250 Hz may be corrected by extending the cochlear filters to include lower CF's. The difference at 2- and 3-kHz may also be narrowed by extending the cochlear filter set to include filters of higher CF's as was done for EIH-LM3.

The better match between the predicted loudness level function for 3 kHz and the observed loudness level function for 8 kHz seems to confirm the suggestion that the loudness of a tone at a frequency close to the upper limit of frequency range grows slower relative to a 1-kHz tone because there are fewer cochlear filters available for excitation at a place on the basilar membrane close to the base of the cochlear partition. In the EIH-LM3 model, due to the phase-locking property simulated by the model, 3 kHz is near the upper frequency range of the model (4 kHz is the maximum frequency).

5.1.3 Iso-loudness Contours

Fig. 5.4 shows observed and predicted iso-loudness contours for pure tones over a frequency range of 150 Hz to 4 kHz and a sound pressure range of 10 to 100 dB SPL. In panel (a), the iso-loudness contours predicted by EIH-LM1 and EIH-LM2 are compared to the observed

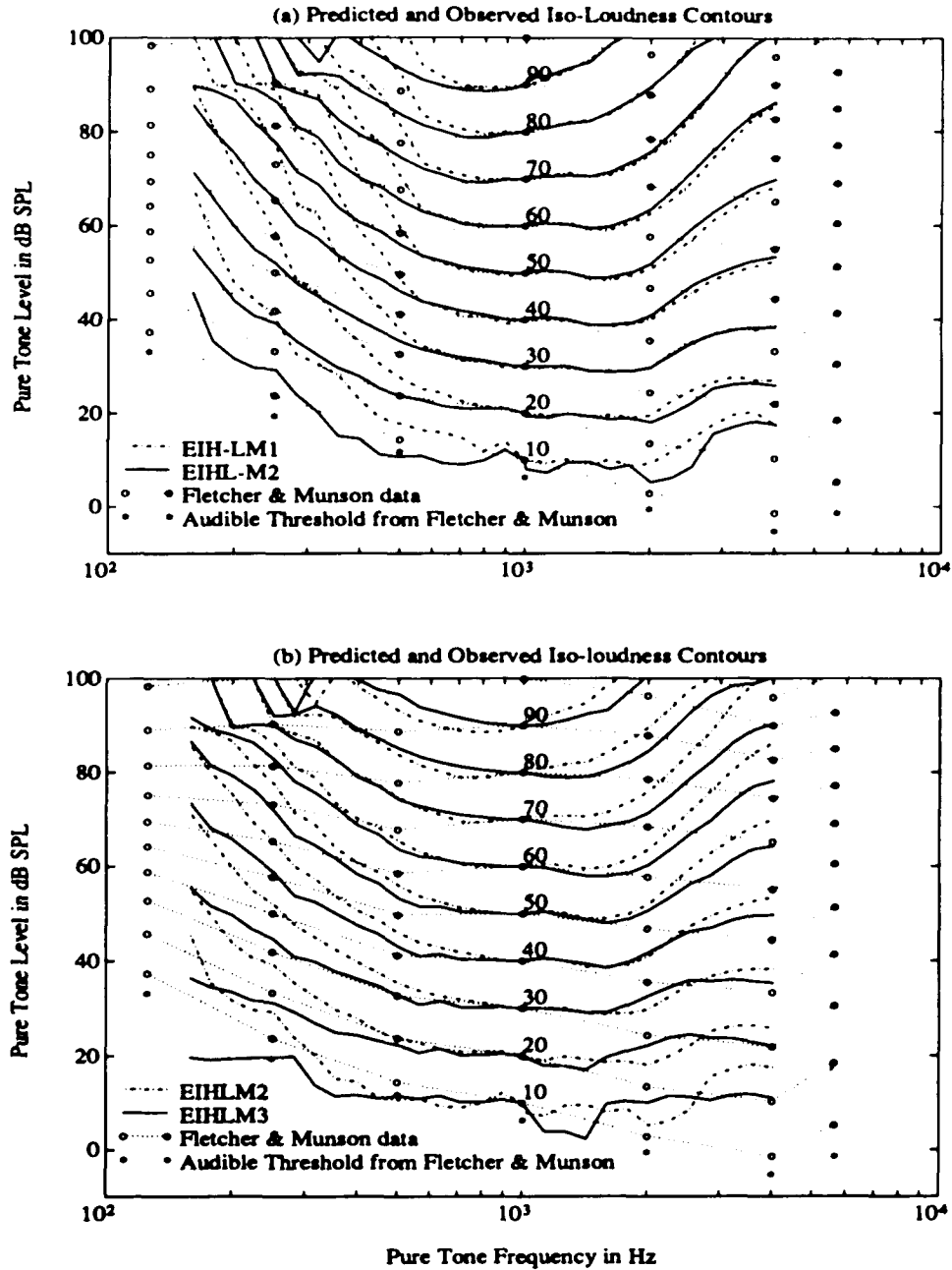


Figure 5.4: Comparison of the Observed and Predicted Iso-loudness Contours for Pure Tones. Comparison of the pure-tone iso-loudness contours predicted by EIH-LM1, EIH-LM2, and EIH-LM3 with the iso-loudness contours derived from the loudness level functions measured by Fletcher and Munson (1933). The top panel contains the iso-loudness contours predicted by EIH-LM1 and EIH-LM2. The bottom panel contains the iso-loudness contours predicted by EIH-LM2 and EIH-LM3. In both panels, the open circles represent the loudness levels at 10 to 90 phons interpolated from the loudness level functions measured by Fletcher and Munson. The open circles joined by dotted lines simulate the iso-loudness contours derived by Fletcher and Munson. The asterisks represent the pure tone thresholds measured by them. In each panel, the abscissa represents frequencies of the pure tones and the ordinate represents the sound pressure levels of the reference tone in dB SPL. The parameter of each curve is the loudness level in phons.

iso-loudness contours. In panel (b), the iso-loudness contours predicted by EIH-LM2 and EIH-LM3 are compared to the same set of observed iso-loudness contours. The observed iso-loudness contours are derived using the loudness level functions measured by Fletcher and Munson (1933) according to the following steps, similar to the way Fletcher and Munson derived theirs. Firstly, for a given frequency, the loudness levels at 10 to 90 phons are interpolated from the loudness level functions measured by Fletcher and Munson. The interpolated loudness levels are represented by the open circles in both panels. Secondly, the derived loudness levels (open circles) are joined by the dotted lines instead of being fitted to curves as Fletcher and Munson did theirs. All iso-loudness functions, predicted or observed, are specified relative to the sound pressure level of an equally loud 1-kHz reference tone. Also shown is the threshold of audibility curve obtained by Fletcher and Munson. In each panel, the abscissa represents frequencies of the pure tones and the ordinate represents the sound pressure levels of the reference tone in dB. The parameter of each curve is the loudness level in phons.

The observed iso-loudness contours of Fletcher and Munson were obtained for binaural listening using head receivers (headphones), while the predicted iso-loudness contours were computed for a monaural listening condition with the SPL of the test and reference sounds specified at the eardrum. Comparing monaural loudness level data to binaural loudness level data does not present a problem, as demonstrated by Fletcher and Munson (1933). They showed that two tones observed to be equally loud binaurally were also judged to be equal in loudness monaurally. Therefore, the predicted iso-loudness curves for the various EIH-LM models apply equally well to monaural and binaural loudness level data. The observed threshold of audibility curves, represented by the asterisks, were measured

monaurally using headphones.

Comparing the iso-loudness contours predicted by EIH-LM1 and EIH-LM2 to the observed iso-loudness contours shows that below about 500 Hz, the iso-loudness contours predicted by EIH-LM1 bend sharply up, indicating the loudness growth of tones in this frequency region decreases sharply relative to the loudness growth of the 1-kHz reference tone. The lower the frequency, the slower the loudness growth rate of the test tone. The iso-loudness contours predicted by EIH-LM2 are much closer to the observed contours below 500 Hz. However, the slight compression of the observed iso-loudness contours at low frequency regions at low to mid SPL is not present in the iso-loudness contours predicted by EIH-LM2. At high SPL, e.g. at 90 phon, flattening of the observed iso-loudness contours, is not present in the predicted contours. At high frequencies, the predicted iso-loudness contours curve upward away from the observed contours.

In panel (b), the iso-loudness contours predicted by EIH-LM3 match the observed iso-loudness contours better than those predicted by EIH-LM2, especially at low frequencies below 60 dB phon and at high frequency at mid to high SPL.

5.1.4 Discussion

A comparison between the observed and the predicted iso-loudness contours in panel (a) demonstrates the improvement of EIH-LM2 over EIH-LM1, and a comparison in panel (b) demonstrate the improvement of EIH-LM3 over EIH-LM2.

The removal of the 40-msec temporal integration time in EIH-LM2 drastically improved the iso-loudness contours predicted by EIH-LM1 below 500 Hz. It has been explained in previous sections and in Chapter 4 that the 40-msec limit causes pure tones of frequency lower than 500 Hz to have fewer than 20 intervals included in the excitation pattern for

loudness summation. The number of intervals used for loudness summation decreases with the frequency of the tone. In comparison to the tones of frequency higher than 500 Hz, where 20 excitations are used for loudness summation, the predicted loudness for tones below 500 Hz is much smaller. Hence the EIH-LM1 contours rise rapidly below 500 Hz. This further supports the notion discussed in Section 5.1.1 that the integration time is related to frequency, i.e. the integration time is determined by the number of cycles needed to establish the loudness.

The extension of the cochlear set to include cochlear filters with higher characteristic frequencies (CF) improves the predicted iso-loudness contours for both low and high frequency regions. At the low frequency region, the improvement is greater for low intensities. At the high frequency region, the improvement is greater for high intensities. However, in both regions, the predicted iso-loudness contours fall short of the observed iso-loudness contours.

At low frequencies, the loudness growth rates are too slow from mid to high SPL. The problem is, as discussed earlier, due to the lack of cochlear filters with CF lower than 170 Hz. Since the cochlear filters are obtained from the cat tuning curves, the cochlear filters lower than 170 Hz are not available. An extension of the filter set to include lower-CF filters should increase the loudness growth rate of low-frequency tones at mid to high SPL.

Due to the lack of cochlear filters of CF lower than 170 Hz, loudness summation for a tone below 200 Hz depends mainly on the low-frequency tails of filters with higher CF. These tails have been extrapolated from the low-frequency portion of the tuning curves. The fact that the predicted 10-phon contour has a threshold lower than the observed data may be an artifact of the extrapolation procedure because several filters with the lowest

CFs have very shallow low-frequency slope leading to those filters being activated at rather low SPL. Thus the slopes of the low-frequency tails and the bandwidths of the cochlear filters, especially those with characteristic frequencies between 150 and 250 Hz, are very important for proper determination of the thresholds and loudness growth rates of tones at low frequency and at low sound pressure levels.

In order to obtain a better prediction and a better understanding of the loudness summation at very low frequency, a more detailed understanding of human cochlear filters in the low frequency region would be important. A method proposed to derive a set of approximated human cochlear filters is presented in Section 8.5.

At the high frequency and high intensity region, the predicted iso-loudness contours are brought closer to the observed contours. This is accomplished by having those additional filters with CF above 7.2 kHz, including the tails of filters whose CF's are above 10 kHz. These additional filters allow excitation pattern to continue growing at high intensity levels. If one considers the frequency limits of the human auditory system, and the way that the iso-loudness contours curve upward at 8 kHz, then perhaps the curving up at the high end in the predicted iso-loudness contours might be expected.

5.2 Predictions for Four-tone Complex

In this section the loudness data of Scharf (1959) using a four-tone complex to examine the critical bands near thresholds are predicted using EIH-LM3. Before Scharf's experiment, Zwicker, Flottorp, and Stevens (1957) showed that keeping the sound pressure constant, the loudness of a noise band with varying bandwidth remained constant until the bandwidth exceeded critical bandwidth. Then the loudness of noise started to increase with increasing

bandwidth for medium and high SPL, but started to decrease with increasing bandwidth for low SPL. The purpose of Scharf's experiment was to measure the critical bands for low- and near-threshold SPL.

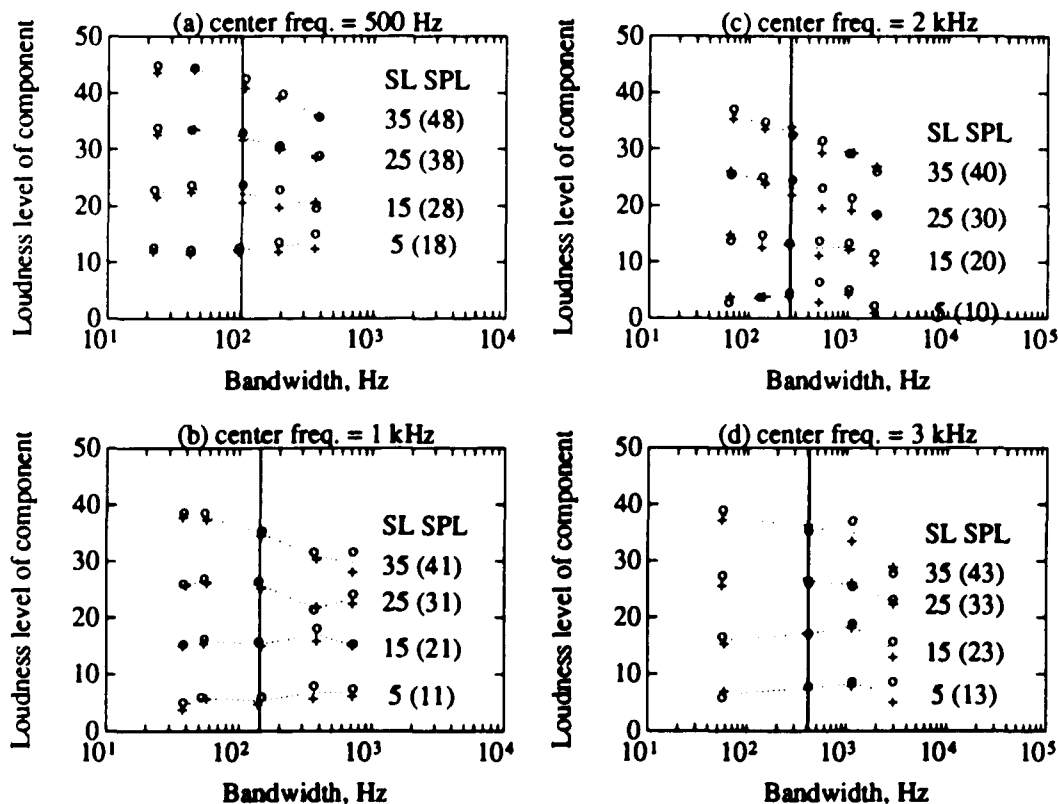


Figure 5.5: Re-plot of Scharf's Critical Band Data (1959, Fig. 3) Measured by Matching the Loudness of Four-tone Complex of Varying Bandwidth to a Comparison Tone at the Arithmetically Centered Frequency of the Complex. The four components of the complex sounds are equal in loudness. The loudness level of each component is specified relative to the comparison tone, i.e. the comparison tone is adjusted in level so as to be equal in loudness to one of the component tones. The SPL of the comparison tone at equal loudness specifies the loudness level of the component tone. The loudness matching data are plotted for the complex as a function of the bandwidth and sensation levels (SL). The four center frequencies are the parameters for the four panels. The open circles (o) represent the medians of the loudness matches obtained by adjusting the level of the four-tone complex, and the plus signs (+) represent those obtained by adjusting the level of the comparison tone. The dotted lines connect the means of all pairs of the matching results, each represented by one open circle and one plus sign. Each solid vertical line indicates Scharf's hypothetical bandwidth for the center frequency.

In Scharf's experiment, the level of a comparison tone was adjusted to match the loudness of a four-tone complex of various bandwidths at four sensation levels (5, 15,

25, and 35 dB SL) and at four center frequencies (500, 1000, 2000, and 3000 Hz). The bandwidth and the number of bandwidths tested depended on the center frequency. The frequency of the comparison tone was equal to the arithmetic center frequency of the complex. The four components of each complex were adjusted to be equal in loudness. The loudness level of each component was obtained by adjusting the level of the comparison tone until it was equal in loudness to the component tone. The SPL of comparison tone at equal loudness defined the loudness level of the component tone. At each combination of sensation level, center frequency, and bandwidth, in addition to adjusting the level of the comparison tone to match the loudness of the four-tone complex, the four-tone complex was also adjusted to match the loudness of the comparison tone. The measured loudness levels of the components are plotted as a function of the bandwidths of the four-tone complex in Fig. 5.5. The way Scharf plotted his data is similar to Fig. 5.5, except in his plot, each dotted line is replaced by a set of two hypothetical line segments with a break at the frequency believed to be the critical bandwidth at the center frequency, i.e. the critical bandwidth measured by Zwicker, Flottorp, and Stevens.

In Fig. 5.5, the center frequencies are the panel parameter and the sensation levels are the curve parameter in each panel. The open circles represent the median loudness levels obtained by adjusting the level of the four-tone complex in the loudness matching task and the plus signs, +, represent the median loudness levels obtained by adjusting the comparison tone in the loudness matching task. The dotted lines connect the means of each loudness-matching pair (one open circle and one plus sign). The solid vertical line in each panel indicates the bandwidth of the critical band estimated by Scharf for the four center frequencies.

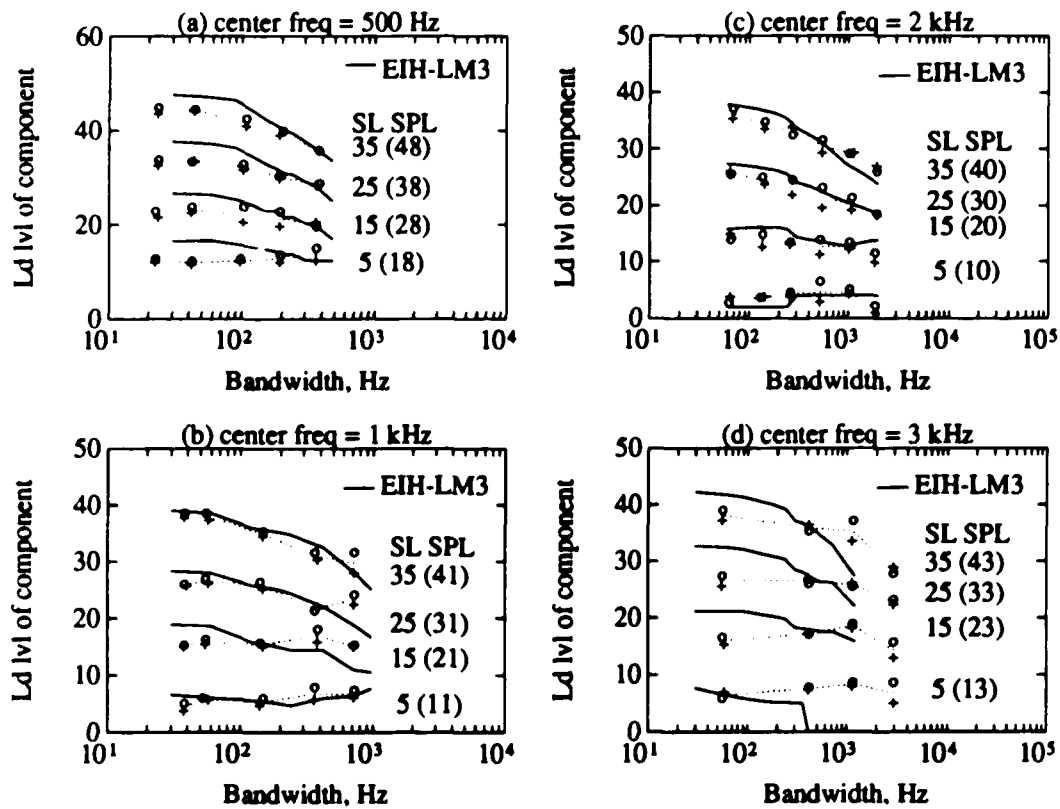


Figure 5.6: Predicted and Observed Loudness Levels for Four-tone Complexes. The solid lines show predicted loudness levels using the EIHL-M3 model. The data, represented by dotted line are from the study by Scharf (1959) on the measurement of critical bands using four-tone complexes. The format of the diagram is the same as for Fig. 5.5.

Model Prediction Predictions of the EIH-LM3 model for the loudness level of four-tone complexes are shown in Fig. 5.6. The predicted loudness levels are represented by solid lines, the observed loudness levels by symbols joined by the dotted lines. The format of the figure is the same as for Fig. 5.5. The EIH-LM3 model computed the loudness levels of the four-tone complex based on the SPL of the tone complex, as used in Scharf's experiment, and not on the sensation level relative to the EIH-LM3 thresholds for the comparison tone or the tone complexes, because the predicted and observed auditory thresholds do not match exactly, especially at low and high frequencies. That is, depending on the center frequency, 18 - 48, 11 - 41, 10 - 40, or 13 - 43 dB SPL are used instead of 5 - 35 dB SL. The predicted thresholds for pure tones around 2 and 3 kHz are elevated because the overall envelope of the cochlear filter set reflecting the middle-ear transfer function is elevated in this frequency region (See Fig. 4.6.) This elevation in turn causes the thresholds of the level detectors to be elevated, as shown in Fig. 4.3. The pure tone thresholds of the EIH-LM model are determined by the threshold of the lowest level detector in each channel.

The predicted and observed loudness levels show relatively good agreement for tone complexes centered on 1 and 2 kHz, as shown in panels (b) and (c). The comparisons for center frequencies at 500 Hz in panel (a) and 3 kHz in panel (d) show that for small bandwidths, the predicted loudness levels are much higher than the observed loudness levels. In panel (a), at the 30-Hz bandwidth the difference in levels is as high as 5 dB. In panel (d), at the 60-Hz bandwidth, the difference in levels is as high as 6 dB. The higher loudness level indicates that the predicted loudness is softer than the observed loudness in these regions. As the bandwidth increases, the predicted loudness grows at a higher rate than the observed loudness with increasing bandwidth, hence the predicted loudness level

of the tone component had to decrease to remain equally loud to the comparison tone. In the case of 500-Hz center frequency (panel (a)), the predicted and the observed iso-loudness curves merge, and match reasonably well between 100-Hz and 500-Hz bandwidths. In the case of 3-kHz center frequency (panel (d)), the predicted and observed iso-loudness curves cross each other at about 500-Hz bandwidth. The predictions are relatively poor for almost all of the bandwidths considered.

The good agreement between observed and predicted data at mid frequencies is to be expected because the predicted loudness for the mid-frequency range at low SPL's seems rather close to the observed loudness as shown in previous predictions for various pure tones. At both lower and higher frequency regions, the predicted loudness is not close to the observed loudness. To truly understand how various parameters of the EIH-LM3 model affect the loudness predictions for this set of observed data and whether the critical bandwidth can be observed and determined from the predictions will require further study.

Normalized Critical-band Data When Scharf's data are plotted without the hypothetical line segments, as shown in his figure (1959, Fig. 3), the critical bandwidths estimated by Scharf at these four center frequencies (the solid lines in Fig. 5.5) seem to disappear. There are no clear break points in the curves at the estimated bandwidths. When break points do occur, the point is not at the vertical line (Scharf's estimate). For most of the curves, the loudness levels of the components change with bandwidths continuously rather than abruptly. This implies that, keeping the sound pressure constant, the loudness of the four-tone complex changes with the bandwidths mostly in a continuous manner and not abruptly. Another conclusion by Scharf of the experiment that has been controversial is that the critical bandwidths are independent of the levels of the complex.

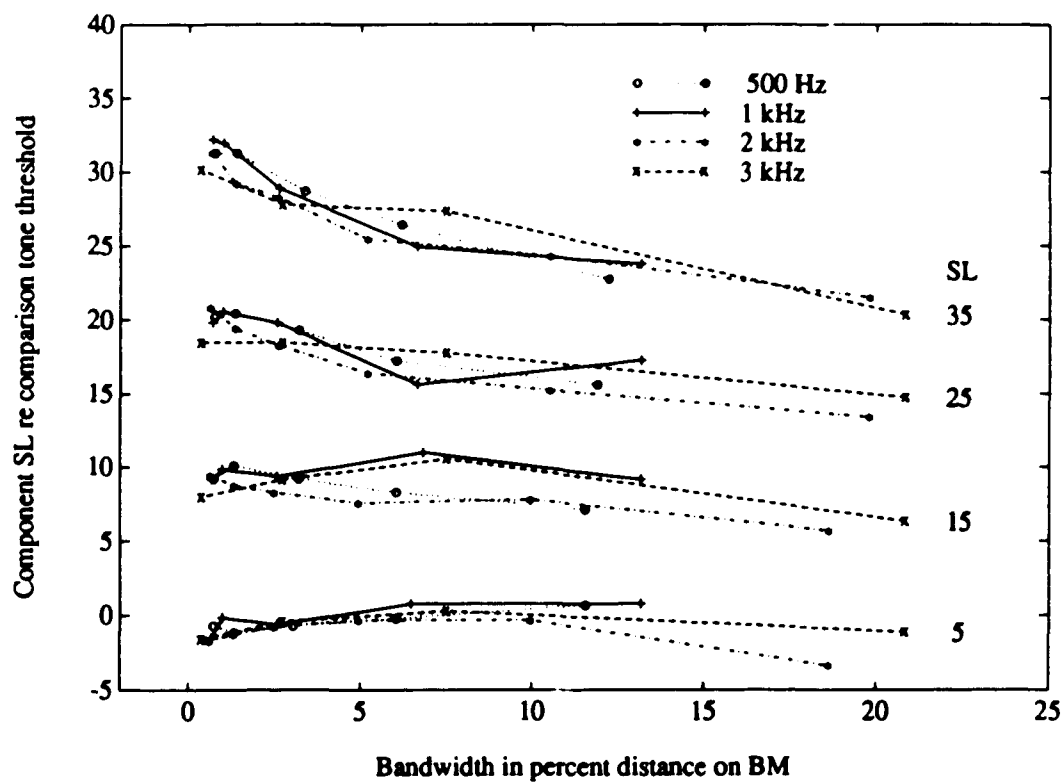


Figure 5.7: An Alternative Representation of Scharf's (1959) Data. Scharf's critical band data (1959, Fig.3) are re-plotted after converting the bandwidths in frequency to bandwidths in percent distance on basilar membrane (BM), and converting the loudness levels of the component tone (re SPL of comparison tone) to the sensation levels (re comparison tone). The converted functions for the four center frequencies are shown on the same diagram (rather than in separate panels, as in Figures 5.5 and 5.6) in order to demonstrate the uniformity of the physiological units in the auditory periphery.

Based on the data shown in Fig. 5.5, the conclusion does not seem to hold. For an example, in panel (b), if one attempted to identify break points on these curves, the estimates would differ from one curve to the next. It has long been believed that the critical-bands describe some physiological principle underlying the psychophysical observations. But the aforementioned conclusions by Scharf requires further study.

Although Scharf's method of estimating the critical band from his data is subject to criticism, the experimental data are nevertheless very useful. We believe there are critical bands in the auditory system and that these critical bands are evident in various psychophysical experiments. In the four-tone experiment, the change in loudness with increasing bandwidth may be gradual rather than abrupt, and if there is a measurable critical band, the bandwidth appears to be level dependent.

In order to investigate the physiological mechanisms underlying the psychophysical observations of critical bands, the four-tone loudness level data obtained by Scharf are re-plotted in place coordinates on the basilar membrane (BM) and in sensation level instead of loudness level (re SPL of the comparison tone). To convert the bandwidths from the frequency domain to the place domain on the BM, Greenwood's cochlear frequency-place function for human (Equation 2.1 and Table 2.1) is used, i.e. $x = 1/2.1 \log(f/165.4 + 1.0)$, where f is the frequency and x is the place on BM as a percent distance from the helicotrema.

The sensation levels of the component tones for a complex are computed by subtracting the threshold of the comparison tone from the loudness levels of the component tones for each center frequency. Note that the loudness levels of the component tones are the dB SPL of the comparison tone equally loud to the component tones and that the four tones in a four-tone complex are equally loud. The threshold of each comparison tone is the

difference between the SPL and the corresponding SL in each panel. For example, the threshold for the 500-Hz comparison tone is obtained by subtracting 5 dB SL from 18 dB SPL giving 13 dB SPL in panel (a) of Fig. 5.5. Similarly, the threshold for the 1-kHz comparison tone is obtained by subtracting 5 dB SL from 11 dB SPL giving 6 dB SPL in panel (b), the threshold for the 2-kHz comparison tone is 5 dB SPL in panel (c), and the threshold for 3-kHz comparison tone is 8 dB SPL in panel (d). Then by subtracting the corresponding thresholds from the loudness levels of the component tones (re SPL of the equally loud comparison tone), one obtains the sensation levels of the component tones (re the threshold of the comparison tone.)

In Fig. 5.7, we superimpose all the converted equal loudness curves for the four center frequencies. In general, the converted curves for each SL show good agreement among four center frequencies. The converted curves of four center frequency agrees the best for 5 dB SL (comparison tone). For the other 3 sensation levels (15, 25, and 35 dB of comparison tone), 3-kHz curve appears to diverge consistently in that at the smallest bandwidth, the sensation levels of the component tones centered on 3-kHz are lower than the SL of the component tones centered at other three center frequencies. At the bandwidth of about 7%, the SL of the component centered on 3-kHz are higher than the others. However, the largest difference between curves is no more than 2.5 dB. The agreement seems to confirm the view that the basic physiological units making up the auditory periphery have similar functional characteristics.

Chapter 6

Loudness Matching Experiment

6.1 Introduction

In order to evaluate the performance of the EIH-based loudness model (EIH-LM3) for complex sounds such as speech, an experiment was conducted in which the loudness levels of *speech (SP)*, a *narrow-band noise (NBN)* and *square wave (SQW)* were obtained. The experimental design and the measured loudness levels are described and discussed in this chapter. The model predictions of these data are reported in the next chapter..

The objective of the experiment was to measure the monaural loudness levels of three different stimuli (SP, NBN, and SQW) at three different stimulus levels for three experimental conditions. In the first condition, loudness levels were measured for the test stimuli in quiet. In the second condition, the test stimuli were presented against a background of noise and loudness levels of the partially masked stimuli were measured. In the third condition, the test stimuli were presented with the same background noise, but in this case the overall loudness of the test stimulus plus the background noise was measured. The three conditions are identified as Q for loudness level in quiet, P for *partial* loudness level in noise, i.e. the loudness level of the partially masked stimulus, and O for the *overall*

loudness level of stimulus plus noise, respectively. The background noise was a modified pink noise at a level of 50 dB re the detection threshold of a 1-kHz tone. The modified pink noise has a spectrum shape of the type recommended by Scharf (1961) and is identified as MPN (see Section 6.2.3 for further details) .

Loudness levels were measured, firstly, in the traditional way by matching the loudness of the test stimulus to that of a 1-kHz tone. Loudness levels were then measured using narrow-band noise as the reference stimulus. Since narrow-band noise was also used as one of the test stimuli, different bandwidths were used. The bandwidth of the narrow-band noise used as a test stimulus was 150 Hz, and the bandwidth of the narrow-band noise used as a reference stimulus was 120 Hz. Both noise bands were centered arithmetically at 1 kHz. These stimuli are identified as NBN150 and NBN120, respectively.

The reason for including NBN as a reference stimulus, in addition to the traditional use of a 1-kHz tone as the reference, was that narrow-band noise sounded more like the test stimuli and, as a consequence, loudness matching was expected to be more reliable using NBN as the reference. It was also of interest to compare loudness levels obtained with the two reference stimuli in order to determine (a) whether different reference stimuli would give significantly different results, and (b) whether one reference stimulus would be more reliable than the other as measured by test-retest repeatability.

The levels of all stimuli were specified relative to the subject's threshold of detection for a 1-kHz tone. For example, when a speech signal was presented to a subject at 60 dB, it meant speech level¹ sound pressure level was 60 dB above that subject's threshold of detection for 1-kHz tone. It was not a true sensation level, but sensation level re the

¹The measure of speech level used in this experiment was the 90-percentile-point SPL (90%-point SPL), the SPL that exceeded 90% of the instantaneous SPL.

subject's 1-kHz tone detection threshold. This kind of sensation level is referred to as SL1k. The use of sensation level (SL) rather than sound pressure level (SPL) to specify signal levels was based on the results of Hellman and Zwislocki (1961) who showed smaller inter-subject variations for signals specified in terms of sensation level. Sensation levels were defined relative to the detection thresholds of a 1 kHz tone in order to provide a common reference for the three types of test stimulus (SP, NBN, SQW).

Loudness levels were measured using two adjustment methods. In one adjustment method, the test stimulus was fixed in level, and the reference stimulus was adjusted in level to match the loudness of the test stimulus. In the second adjustment method, the reference stimulus was fixed in level and the test stimulus was adjusted to match the loudness of the reference stimulus. Loudness matches were obtained at three fixed stimulus levels for each experimental condition and method of adjustment, as shown in Table 6.1. Different sets of stimulus levels were used for Q, P, and O conditions because

Stimulus Levels in dB SL1k		
Experimental Conditions	Levels of Test Stimulus (Reference Adjusted)	Levels of Reference Stimulus (Test Stimulus Adjusted)
Quiet (Q)	30, 50, 70;	30, 50, 70;
Partial-loudness (P)	50, 60, 70;	30, 50, 70;
Overall-loudness (O)	40, 50, 60;	70, 72, 75;

Table 6.1: Stimulus Levels Used for Each Experimental Condition and Method of Adjustment.

the patterns of loudness growth with level differed among conditions; the stimulus levels were chosen so as to cover similar ranges of loudness level for each experimental condition and to capture the variations in loudness growth patterns over a large range. For the same reasons, two sets of stimulus levels were used for the two adjustment methods for P and O

experimental conditions; where the test stimulus and the reference stimulus were presented in asymmetrical situations. In condition P, the reference stimulus in quiet was matched to the test stimulus partially masked by the background noise, while in condition O, the reference stimulus in quiet was matched to the test stimulus plus the background noise.

6.2 Experimental Design

6.2.1 Procedure

The experiment consisted of six test sessions, with two sessions (A and B) for each of the Q, P, and O conditions. In session A, the *reference* stimulus was adjusted to match the test stimulus in loudness. In session B, the *test* stimulus was adjusted to match the reference stimulus in loudness. The two sets of three stimulus levels used for the test stimulus levels in session A and for the reference stimulus in session B are summarized in Table 6.1. Note that the set of three stimulus levels selected for the reference in session B of condition O were initially 65, 70, and 75 dB SL1k. However, the three stimulus levels were changed to 70, 72, and 75 dB SL1k after three subjects were tested for this session. The change resulted from these subjects' complaints that they were unable to adjust for equal loudness for some of the test conditions because the background noise alone in those test conditions sounded louder than the reference stimulus. Hence, the summation of the background noise plus the test stimulus could not be equal to the reference stimulus in loudness at the given stimulus level. Their data showed that the troubles were with 65 dB SL1k.

A total of twelve normal-hearing subjects (4 females and 8 males) were selected for the experiment (see Appendix A.1 for subject information). Each subject received a different sequence of test sessions according to Latin-square design (see Appendix A.2). Eleven out

of twelve subjects completed all six sessions, while one subject completed five sessions (two sessions for each of conditions Q and O and one session of condition P). Among the 11 subjects who completed all sessions, 3 were tested for session B of condition O using the initial set of the reference stimulus levels (65, 70, and 75 dB SL_{1k}). Data collected from these three subjects for this specific session were incomplete.

In each test session, 54 loudness levels were measured: three replications for each of 3 different test stimuli (SP, NBN, and SQW) at 3 different levels using 2 different reference stimuli (PT and NBN120). The 54 test conditions were presented to the subject in random order. The measurement of loudness level for each test condition is referred to as a *trial*.

The method of adjustment (with bracketing) was used to obtain the loudness level in each trial. Pairs of stimuli were presented cyclically with a period of 5.2 to 5.6 seconds. The first stimulus of a pair was fixed in level and is referred to as the standard. The second stimulus of the pair was adjusted in level to match the loudness of the standard and is referred to as the variable. In session A, the test stimulus was the standard and the reference stimulus was the variable; and vice versa in session B. The subject rotated a knob to adjust stimulus level. When the two stimuli were judged to be equally loud, a button was pressed to proceed to the next trial.

The subjects were given both written and oral instructions of the task to be performed at the beginning of each session. These instructions are reproduced in Appendix A.3. All but one² of the subjects had no experience in loudness matching experiments, so they were

²One of the 12 subjects was experienced in the loudness matching task. This subject had the smallest test-retest variability. This subject's standard deviation, averaged over all test conditions, was significantly less than that for two other subjects, but not significantly different from that for five other subjects. (Note only the statistics of eight subjects, whose data were complete for the whole experiment, was compared.) This subject's loudness growth patterns for different stimuli in different test conditions were not much different from others'. The loudness level averaged over all test conditions for this subject was the highest of all subjects, but was only 1.05 standard deviation from the grand mean of all subjects. This difference

reminded before each session to adjust for loudness and not for similarity or annoyance of the signals.

No time limit was set for the loudness matching task, but in order to avoid extremely long sessions (which could result in fatigue or possible variations in judgmental criteria) subjects were encouraged not to prolong the session and to maintain a good overall pace. The duration of each trial, as well as overall session duration, was monitored for later analysis. Only one session was performed on a single day. The time taken to complete all six sessions ranged from 8 days to six weeks.

6.2.2 Equipment

The experiment was automated using a Masscomp 5400 computer. The block diagram of the equipment is shown in Figure 6.1.

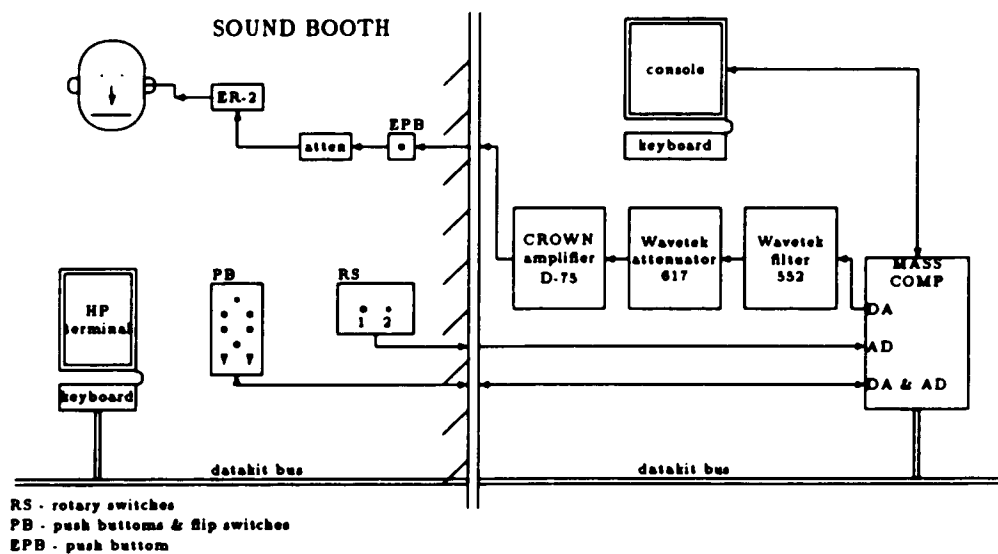


Figure 6.1: The test is administered by the Masscomp computer. The subject in the sound booth communicates to the computer via an HP-terminal, a knob box (RS), and two push-button boxes (PB and EPB).

was not statistically significant.

The computer generated the stimuli and recorded the data. The Wavetek filter (model 552) low-passed the signal to 7.2 kHz. The signals were delivered to the subject's ear canal via an ER-2 TubePhone insert earphone (Etymotic Research) and a disposable foam ear-tip. The subject used the 10-turn knob RS to adjust the intensity of the stimulus. The knob controlled a 12-bit A/D converter and the dynamic range of adjustment was controlled by the computer. Upon completing a trial, the subject pushed the button PB to signal the computer to record the signal intensities, trial duration, and related variables and to proceed to the next trial. The push button EPB switched off the stimulus instantly. It was designed for emergency use in case the stimulus was undesirably loud. No subject ever used it. The HP terminal (Hewlett-Packard 2306) was used for subject/computer communication.

Calibration The equipment was calibrated using a 1-kHz pure tone and a dB-100 (Zwislocki) coupler as suggested in the manufacturer's specification for the Etymotic ER-2 TubePhone. According to the manufacturer, this calibration corresponds to the pressure at the eardrum for the average real ear. The equipment was calibrated daily using a system identification program SYSid^{TM3} to ensure the input to the ER-2 TubePhone was at the desired level. We did not perform real-ear measurements at the eardrum. However, measurements of pure-tone thresholds were repeatable and consistent with the manufacturer's specifications, which provides evidence that the calibration procedure was valid. Concerning the effect of the system on broad-band stimuli, we measured the frequency response of the system and of the ER-2 earphone. The frequency response of ER-2 earphone agreed with the manufacturer's specifications. It is flat within ± 1 dB from 200 Hz to 10 kHz. Below

³SYSid is a trademark of the Ariel Corp., Highland Park, NJ.

200 Hz, it decreases at about 2 dB per octave. The frequency response of the equipment at the input to ER-2 earphone is also flat within $+1/-3$ dB from 40 Hz to 7.2 kHz.

6.2.3 Stimuli

Six types of stimuli were used in the experiment: The speech signal (SP) was a recording of a male speaker producing the sentence "He ordered a peach pie with ice cream." The signal was digitized at a sampling rate of 20 kHz with 16 bit precision. The narrow-band noise stimuli (NBN150 & NBN120) were generated by computer and stored in the same way as the speech signal. Both of the NBN stimuli had bandwidths less than the corresponding critical bandwidth, 160 Hz at 1 kHz as specified in Zwicker's (1961) critical-bandwidth table. A square wave (SQW) with a period of 51.2 msec was generated by computer and stored in the same way as the other stimuli. A pure tone (PT) of 1 kHz was generated by computer and stored in the same way as the other stimuli. The background noise was a modified pink noise (MPN) with frequency spectrum shaped according to Scharf's (1961) description, i.e. flat out to 500 Hz, then sloped downward at 3 dB/oct until 8 kHz, followed by a steep downward slope of 48 dB/oct out to 10 kHz. This spectrum shape is meant to provide approximately equal noise power per critical band up to 8 kHz.

Each of the above stimuli was limited in duration to 2.3 seconds except for the masking noise (MPN) in the Partial-loudness Condition, P, where the background noise had a lead and a lag of 200 msec before and after the test signal. The lead and lag times were included in order to assist the subject in distinguishing the test signals from the background noise. All signals were switched on and off with a ramp of 20 msec duration. As mentioned in Section 6.2.1, in each trial, standard and variable stimuli were presented alternately until the subject signaled the computer to proceed to the next trial. The time interval between

the standard and the variable stimuli was 200 msec and between every pair was 400 to 600 msec.

After the collection of the loudness matching data, it was learned that Debout (1959) showed that the threshold of a tone measured after the cessation of noise had small post-stimulatory threshold shift, independent of the noise sound pressure level. Hellman (1970) further showed that a shift of 3 dB at detection threshold was sufficient to alter a masked loudness-level curve below 40 dB SPL. Hellman also found that at an interval of 800 msec, there was no observable effect of noise on threshold. In the case of the loudness matching experiment in this dissertation, the short time interval of 200 msec between the two stimuli of each pair might affect the loudness-level functions measured. However, the lowest stimulus level was about 38 dB SPL, which was close to the level above which the loudness-level curve would not be affected by post-stimulatory threshold shift, as reported by Hellman. Hence it is not clear how much the 200-msec interval between the standard and the variable would affect the loudness level functions obtained in this dissertation.

6.3 Results

Figures 6.2 through 6.7 show the measured loudness levels as a function of test stimulus level (specified in dB sensation level re subject's hearing threshold for a 1-kHz pure tone) for the six test sessions. Fig. 6.2 shows loudness levels for session A of the in-quiet condition (Q) where the *reference* stimulus was adjusted in level. Fig. 6.3 show the corresponding data for session B of the in-quiet condition where the *test* stimulus was adjusted in level. Fig. 6.4 and Fig. 6.5 show the corresponding sets of data for the Partial-loudness condition (P), and Fig. 6.6 and Fig. 6.7 show the corresponding sets of data for the Overall-loudness

condition (O).

Each of the figures consists of six plots. The rows correspond to different reference stimuli. For the upper row of plots, the 1-kHz pure tone (PT) was used as the reference. For the lower row of plots, narrow-band noise (NBN120) was used as the reference. The columns correspond to different test stimuli. For the left-most column of plots, the test stimulus used was speech (SP). For the middle column of plots, the test stimulus used was narrow-band noise (NBN150), and for the right-most column of plots, the test stimulus used was square wave (SQW).

In all of the above figures, each measured loudness level is represented by a single dot. Since there were three observations (replications) for each test condition for each of 9 to 12 subjects (as explained in Section 6.2.1), a total of 27 to 36 dots is shown for each test

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

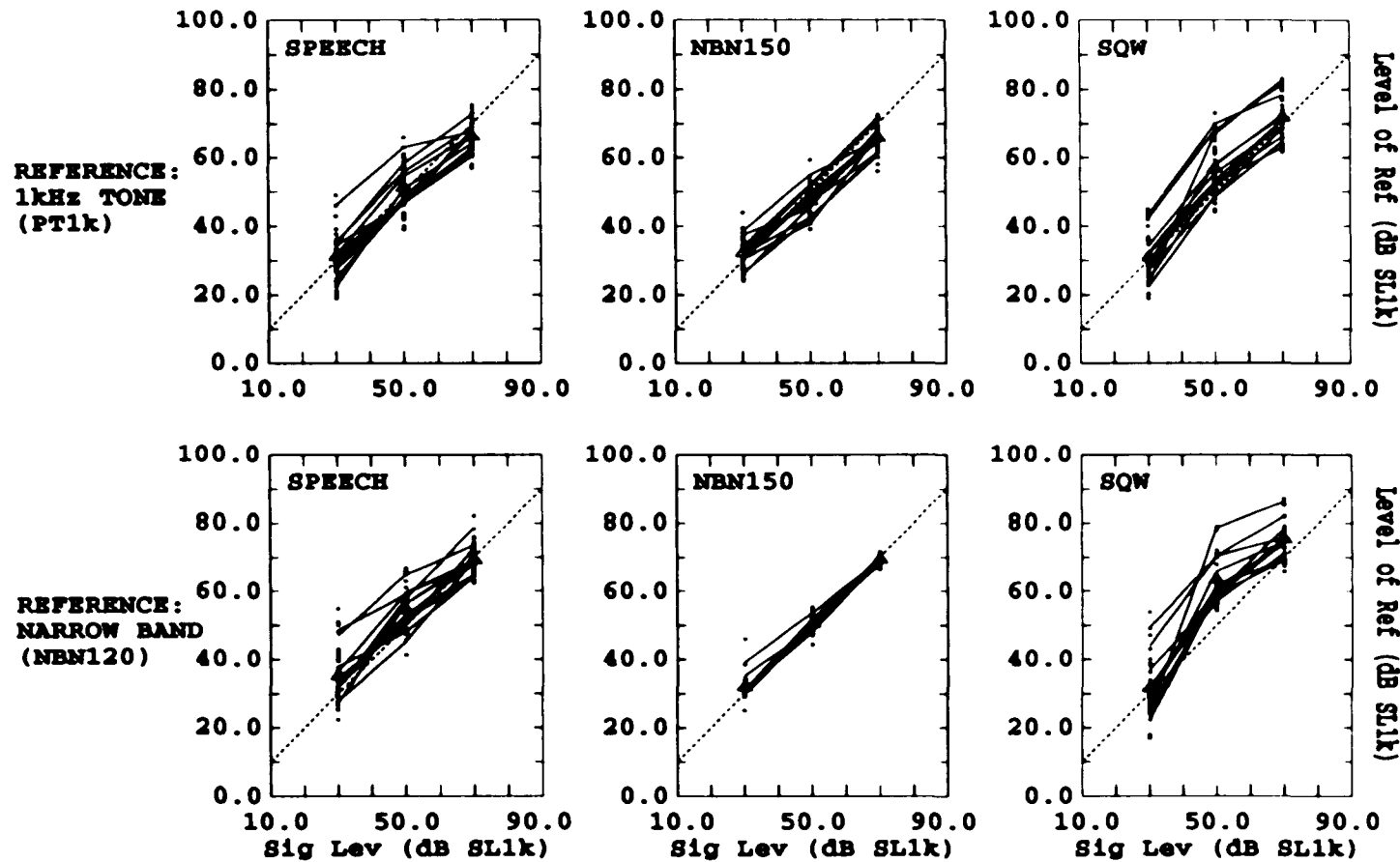


Figure 6.2: Measured Loudness Levels as a Function of Stimulus Levels in SL1k (re 1-kHz Tone Threshold) for the *Quiet Condition (Q)* Where the Reference Stimulus Was Adjusted in Level. For the upper row of plots, the 1-kHz pure tone (PT) was used as the reference. For the lower row of plots, narrow-band noise (NBN120) was used as the reference. For the left-most column of plots, the test stimulus used was speech (SP). For the middle column of plots, the test stimulus used was narrow-band noise (NBN150) and for the right-most column of plots, the test stimulus used was square wave (SQW).

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

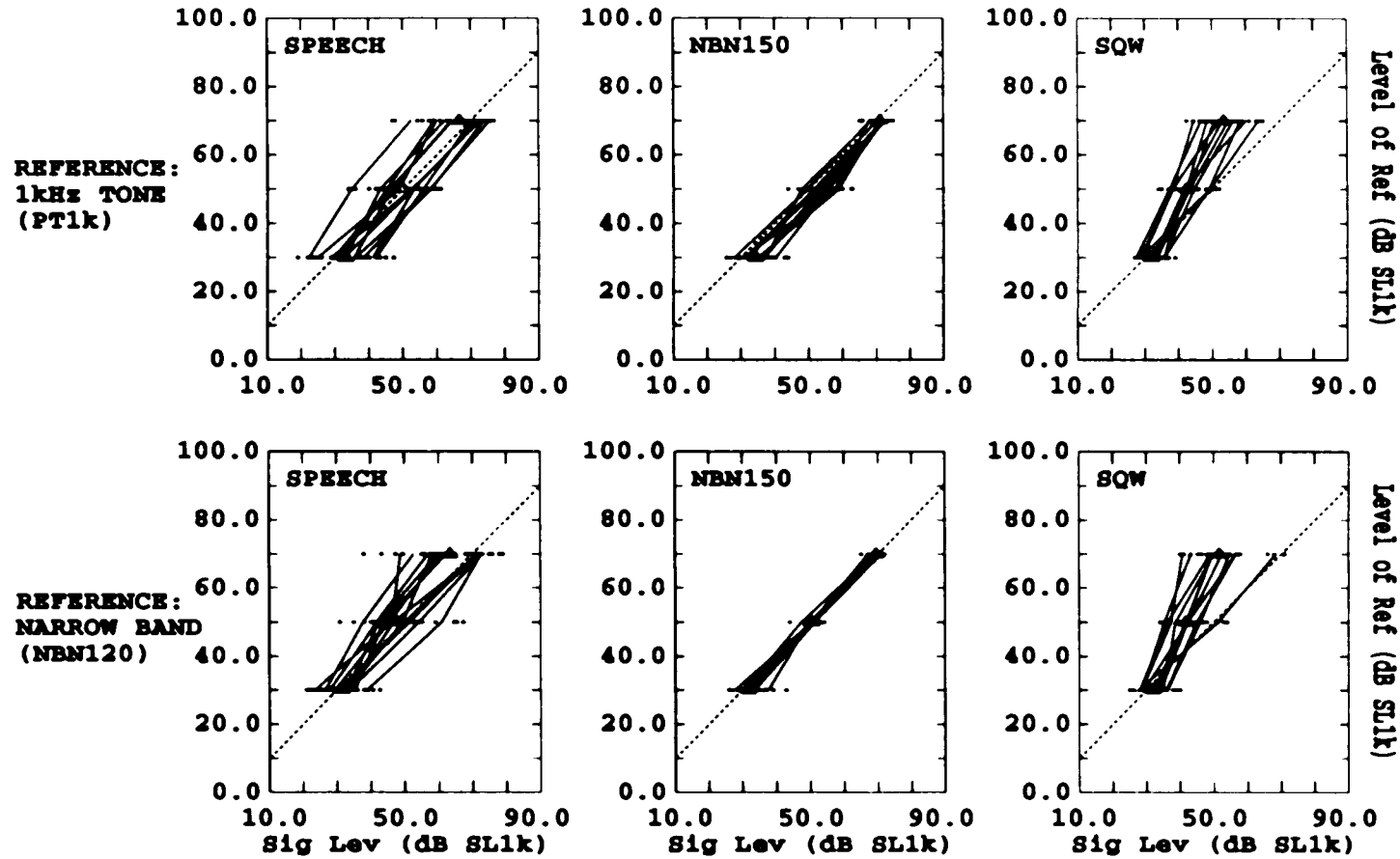


Figure 6.3: Measured Loudness Levels as a Function of Stimulus Levels in SL1k for the Quiet Condition Where the Test Stimulus Was Adjusted. The format of the figure is the same as Fig. 6.2.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

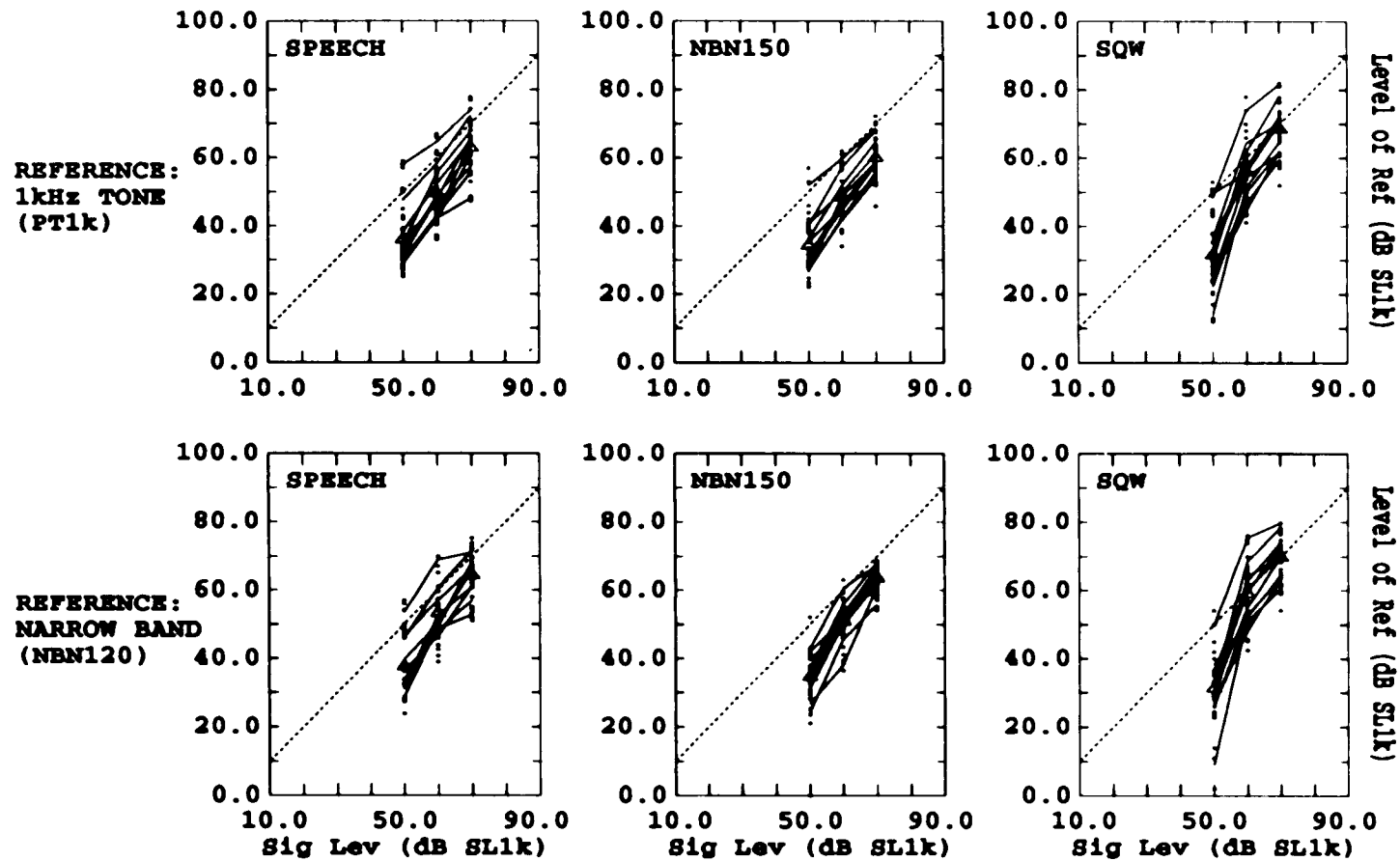


Figure 6.4: Measured Partial Loudness Levels as a Function of Stimulus Levels in SL1k for the Partial-loudness Condition (P) Where the Reference Stimulus Was Adjusted. The format of the figure is the same as Fig. 6.2.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

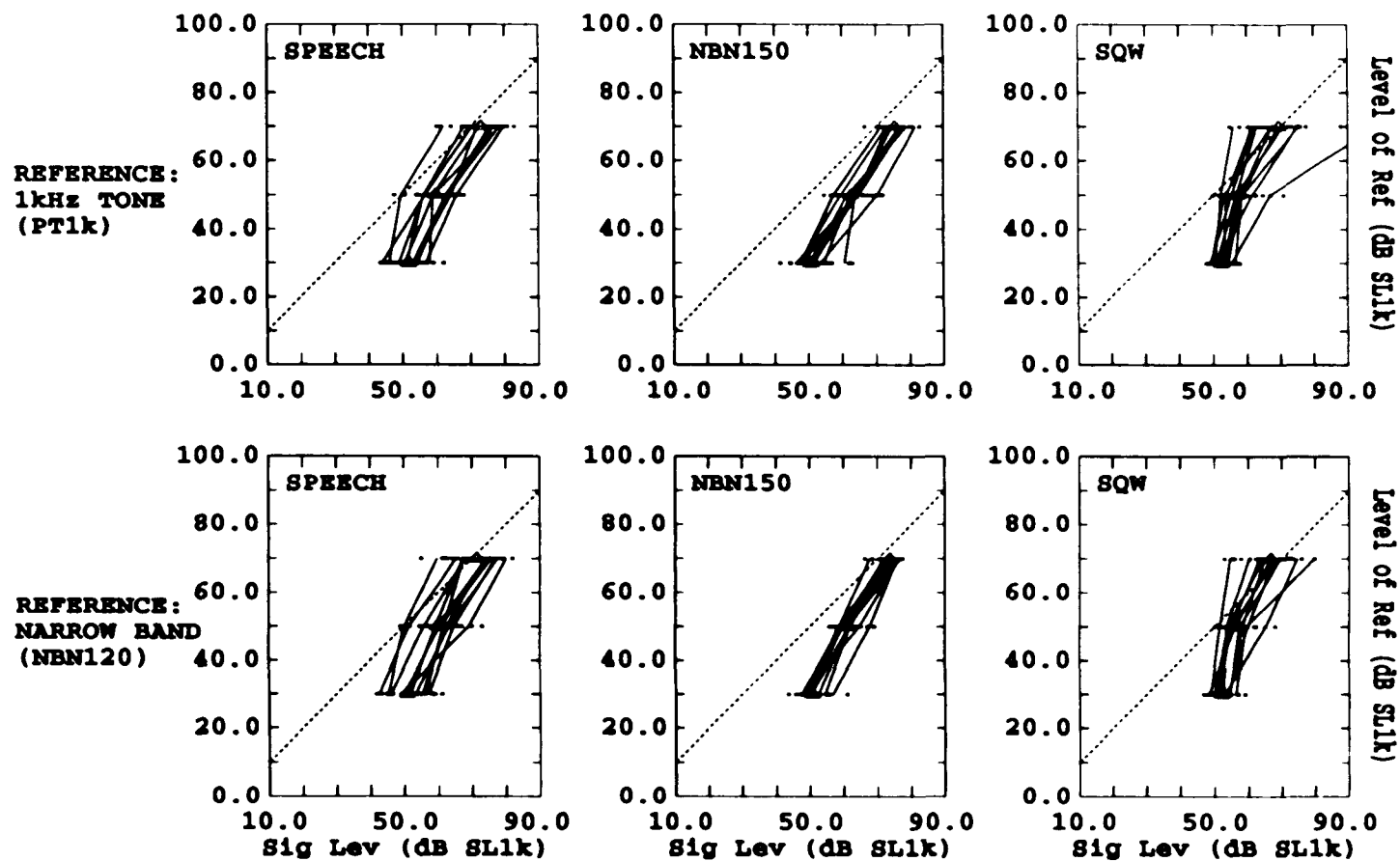


Figure 6.5: Measured Partial Loudness Levels as a Function of Stimulus Levels in SL1k Where the Test Stimulus Was Adjusted. The format of the figure is the same as Fig. 6.2.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

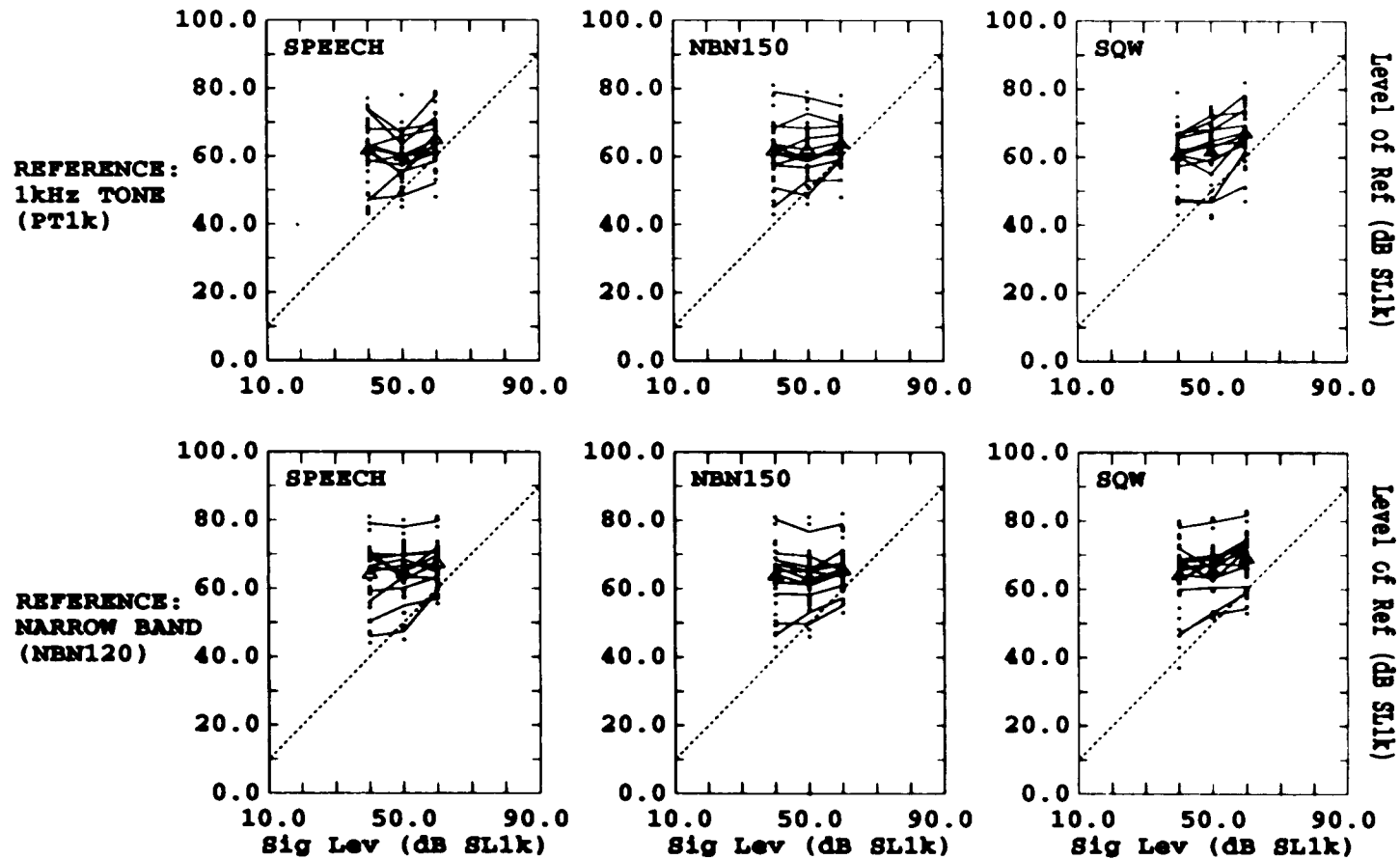


Figure 6.6: Measured Loudness Levels as a Function of Stimulus Levels in SL1k for the Overall-loudness Condition (O) Where The Reference Stimulus Was Adjusted. The format of the figure is the same as Fig. 6.2.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

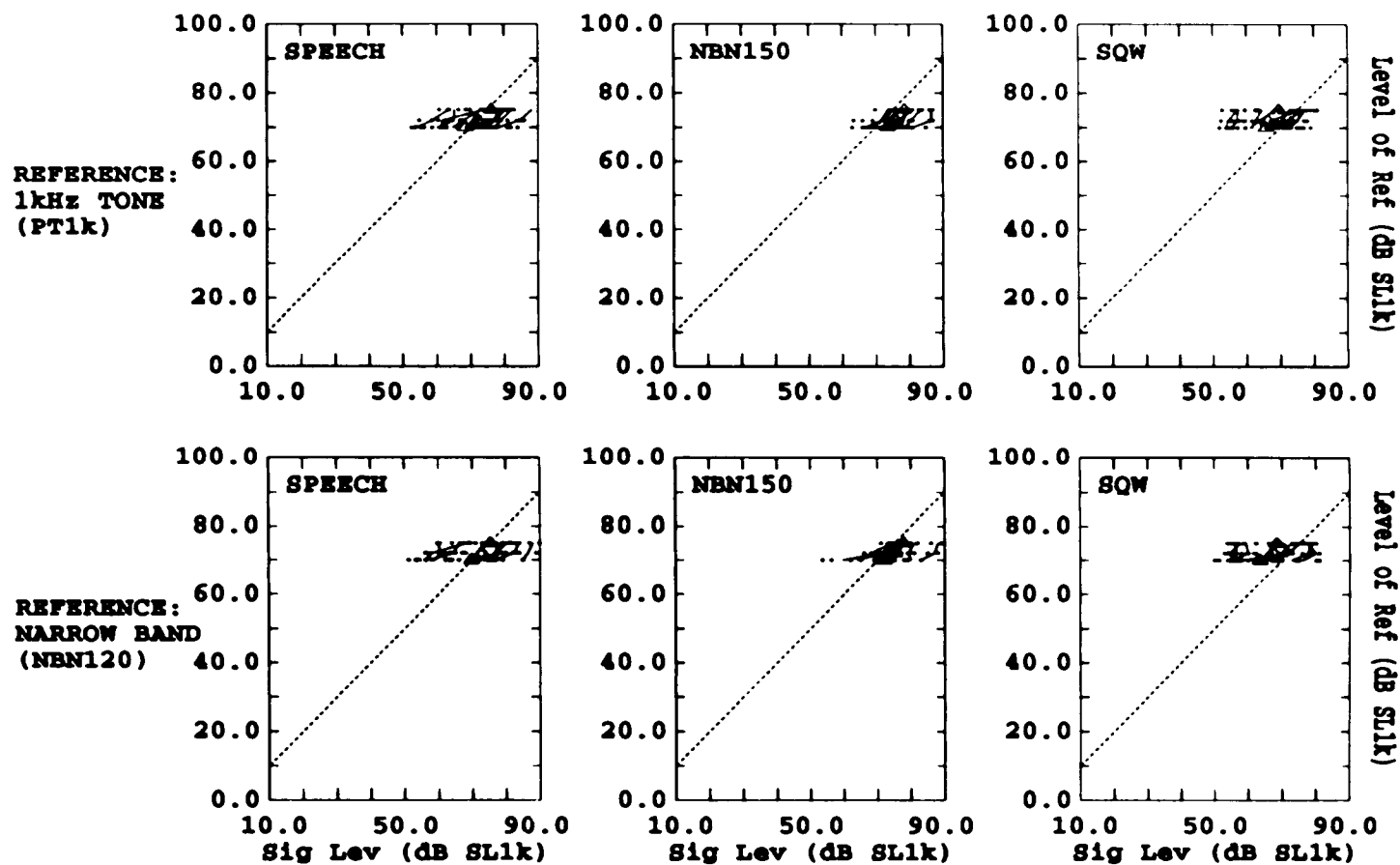


Figure 6.7: Measured Overall-loudness Levels as a Function of Stimulus Levels in SL1k Where the Test Stimulus Was Adjusted. The format of the figure is the same as Fig. 6.2.

condition. The means of the three observations at each of the three test stimulus levels are connected by a solid line which represents the mean loudness growth for each subject. The symbol, Δ , represents the overall mean loudness level averaged across all subjects' mean loudness levels for a given stimulus level. The dashed diagonal line in each panel serves as a guide. Any point on the dashed line indicates that the test and the reference stimuli are equally loud at the same sensation levels re subject's detection threshold for 1-kHz tone.

Three aspects of each solid line are of interest, the slope of the line, the location of the line relative to the dashed line, and the shape of the line. If a solid line is parallel to the dashed line, the subject's rate of loudness growth for the test stimulus is the same as that for the reference stimulus. If the slope of the solid line is steeper than that of the dashed line, the rate of the loudness growth of the test stimulus is greater than that of the reference stimulus (and vice versa, if the slope of the solid line is less than that of the dashed line). If the solid line falls above the dashed line, then the subject requires higher level for the reference stimulus to be equal in loudness to the test stimulus, and vice versa if the solid line falls below the dashed line. The shape of a line indicates a subject's loudness growth pattern for a test stimulus. Since a solid line connects the subject's mean judgements at three signal levels, three shapes are possible, i.e. straight, concave, and convex. A straight line indicates that the subject's loudness growth rates for the two segments (relative to the reference) are the same. A concave line indicates that the loudness of the test stimulus grows at slower rate at lower levels than at higher levels (relative to the reference). A convex line indicates just the opposite.

6.4 Data Analysis

The data show a moderate degree of scatter, largely as a result of inter-subject differences. Although there are large inter-subject differences in the location of each loudness growth curve, the pattern of loudness growth is essentially the same within each experimental condition. These patterns are clearly different for the three experimental conditions, Q, P, and O, but not so clearly different for different test stimuli within each experimental condition.

In order to analyze these loudness growth patterns and the experimental variables affecting these patterns, two analyses of variance (ANOVA) were performed on the data. The first ANOVA examined the test-retest reliability of the subjects' responses. The second ANOVA examined the effects of the experimental variables on the measured loudness levels. The first ANOVA was also used to determine the homogeneity of the test-retest variability across subjects and conditions in order to check the assumptions underlying the subsequent ANOVA.

6.4.1 Test-retest Reliability

The first ANOVA examined whether test-retest reliability differed systematically as a function of the experimental variables. For this analysis, the standard deviation (s.d.) was calculated for each set of three replications⁴, i.e. for each test condition for each subject. A six-way ANOVA with repeated measurements over the subjects was then performed on the measured standard deviations. The six variables (factors) with their values are listed in Table 6.2. The first column specifies the variables by name. The second column identifies

⁴The tables containing the arithmetical group means and standard deviations for various tasks can be found in Appendix A.5.

Experimental Variables			
Variable	Identifier	No. of levels	Variable values
Stimulus Level	L	3	low-, mid-, and high-level
Test Stim. Type	T	3	SP, NBN150, and SQW
Ref. Stim. Type	R	2	PT and NBN120
Adjust. Method	A	2	adj. ref. stim. and adj. test stim.
Exp. Condition	C	3	Q, P, and O
Subject	S	8	S1, S3, S8, S9, S13, S14, S15, & S16

Table 6.2: The first column specifies the variable by name, the second column identifies each of the variables by a capital letter, the third column indicates the number of levels for each variable, and the last column identifies the values of each variable.

each of the variables by a single capital letter. The third column indicates the number of levels for each variable. The last column identifies the values of each variable. The last variable, Subject, is limited to the eight subjects who completed all the tests.

As noted in Section 6.2.1, three subjects (S4, S5, and S7), had missing data for session B of condition O, and one subject, S6, for reasons unrelated to the experiment, did not complete condition P. A subsequent analysis of the limited data sets for each of these subjects showed essentially the same results as the main analysis. For each of these subjects, a five-way ANOVA on the measured standard deviation excluding the condition with missing data was performed. The results for subjects S5, S6, and S7, showed no significant effect at $p \leq 0.002$. The results for S4 showed that the adjustment method (A) was a significant effect. However, this result did not conflict with the main analysis shown in the following table.

The results of the ANOVA are summarized in Table 6.3. Between-subject difference (S) was the only main effect found to be statistically significant ($p \leq .001$)⁵. Two interactions

⁵A low probability ($< .001$) was used since there were 62 simultaneous comparisons, a number large enough for several F-tests to show significance at the 0.05 level purely by chance. Using the Bonferroni rule (Miller, 1966), a safe significant level is $.05/62$ or roughly .001.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

Analysis of Test-retest Variability											
Grand Mean: 2.681552 dB											
Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F - Ratio	Signf. Level	Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F - Ratio	Signf. Level
L	64.60288	2	32.30144	7.407	0.007	LS	61.05203	14	4.36086	1.991	0.035
T	74.26938	2	37.13469	6.911	0.008	TS	75.22833	14	5.37345	2.453	0.009
LT	38.75026	4	9.68756	3.901	0.012	LTS	69.53260	28	2.48331	1.134	0.337
R	39.04902	1	39.04902	14.558	0.007	RS	18.77645	7	2.68235	1.224	0.305
LR	3.34620	2	1.67310	0.509	0.617	LRS	46.05860	14	3.28990	1.502	0.141
TR	13.91882	2	6.95941	2.156	0.151	TRS	45.18119	14	3.22723	1.473	0.152
LTR	4.12823	4	1.03206	0.247	0.908	LTRS	116.97800	28	4.17779	1.907	0.020
A	59.03209	1	59.03209	5.596	0.048	AS	73.81866	7	10.54552	4.814	0.001
LA	30.98538	2	15.49269	4.219	0.036	LAS	51.41412	14	3.67244	1.676	0.087
TA	0.49166	2	0.24583	0.034	0.967	TAS	101.17684	14	7.22692	3.299	0.001
LTA	18.86094	4	4.71523	1.962	0.127	LTAS	67.27718	28	2.40276	1.097	0.375
RA	6.25941	1	6.25941	2.538	0.153	RAS	17.26690	7	2.46670	1.126	0.360
LRA	0.71141	2	0.35571	0.239	0.793	LRAS	20.85361	14	1.48954	0.680	0.784
TRA	0.80601	2	0.40300	0.140	0.871	TRAS	40.43297	14	2.88807	1.318	0.226
LTRA	12.20620	4	3.05155	0.908	0.526	LTRAS	94.09666	28	3.36059	1.534	0.087
C	0.49134	2	0.24567	0.040	0.962	CS	87.03358	14	6.21668	2.838	0.003
LC	6.13125	4	1.53281	0.581	0.682	LCS	73.84479	28	2.63731	1.204	0.272
TC	27.73800	4	6.93450	2.052	0.113	TCS	94.61776	28	3.37921	1.543	0.084
LTC	18.86955	8	2.35869	0.714	0.680	LTCS	185.11426	56	3.30561	1.509	0.063
RC	1.81102	2	0.90551	0.323	0.733	RCS	39.25517	14	2.80394	1.280	0.249
LRC	7.09750	4	1.77438	0.387	0.817	LRCS	128.27136	28	4.58112	2.091	0.010
TRC	38.81709	4	9.70427	2.057	0.113	TRCS	132.09201	28	4.71757	2.154	0.006
LTRC	22.92756	8	2.86594	0.779	0.624	LTRCS	206.05038	56	3.67947	1.680	0.027
AC	47.21902	2	23.60951	5.178	0.020	ACS	63.83817	14	4.55987	2.082	0.027
LAC	52.27334	4	13.06833	5.896	0.002	LACS	62.06244	28	2.21652	1.012	0.472
TAC	14.61951	4	3.65488	1.622	0.196	TACS	63.10714	28	2.25383	1.029	0.451
LTAC	6.29088	8	0.78636	0.213	0.987	LTACS	206.75153	56	3.69199	1.685	0.026
RAC	3.88768	2	1.94384	1.003	0.394	RACS	27.12803	14	1.93772	0.885	0.579
LRAC	7.46688	4	1.86672	0.587	0.677	LRACS	89.00794	28	3.17886	1.451	0.118
TRAC	7.41021	4	1.85255	0.963	0.566	TRACS	52.76104	28	1.88432	0.860	0.662
LTRAC	21.46965	8	2.68371	1.225	0.301	LTRACS	122.67467	56	2.19062		
S	70.53712	7	10.07673	4.600	0.0010	TOTAL	3255.22996	863			

Table 6.3: An asterisk indicates that the effect is significant at $p \leq 0.001$. The six factors considered were L (stimulus level), T (test stimulus type), R (reference stimulus type), A (Adjustment method), C (experimental condition), and S (subject).

AS and TAS, were found to be statistically significant ($p \leq .001$). A post-hoc analysis (Tukey, 1949) for the main effect S showed that subjects could be subdivided into two groups with 5 overlapping members (Group 1 consisting of S1, S16, S8, S3, S15, and S13, and Group 2 with S16, S8, S3, S15, S13, S14, and S9). Subjects within a group did not differ significantly from each other. Thus, of the 8 subjects, only subject 1 (with the lowest test-retest standard deviation of 2.1 dB) differed significantly from subjects S14 and S9 (with the highest test-retest standard deviations of 2.9 and 3.1 dB, respectively).

The interaction AS, as shown in Panel (a) of Fig. 6.8, indicated that, except for S15, all subjects had higher test-retest variability when the reference stimulus was adjusted than when the test stimulus was adjusted. This is true with S4, S5, and S7. For S6, the two methods of adjustment showed equal standard deviations.

The interaction TAS indicated that the AS interaction also varied with Test Stimuli (T). As shown in Panels (b) to (i) of Fig. 6.8, of the 7 similar subjects, S14 (in (g)) showed the opposite effect for NBN150 (test stimulus type 2), and S16 (in (i)) for SP (type 1), although in both cases the differences were very small. The same was with S15 (in (h)), who showed higher test-retest variability when the test stimulus was adjusted, albeit not for all test stimuli. S15's mean s.d.'s of the two adjustment methods for NBN150 were in the same range as other subjects. Among the four subjects not included in the main analysis, as mentioned previously, there was no significant interaction between variables T and A.

The result of this analysis showed that, overall, the standard deviations of the three measurements for each test condition did not vary systematically with the experimental

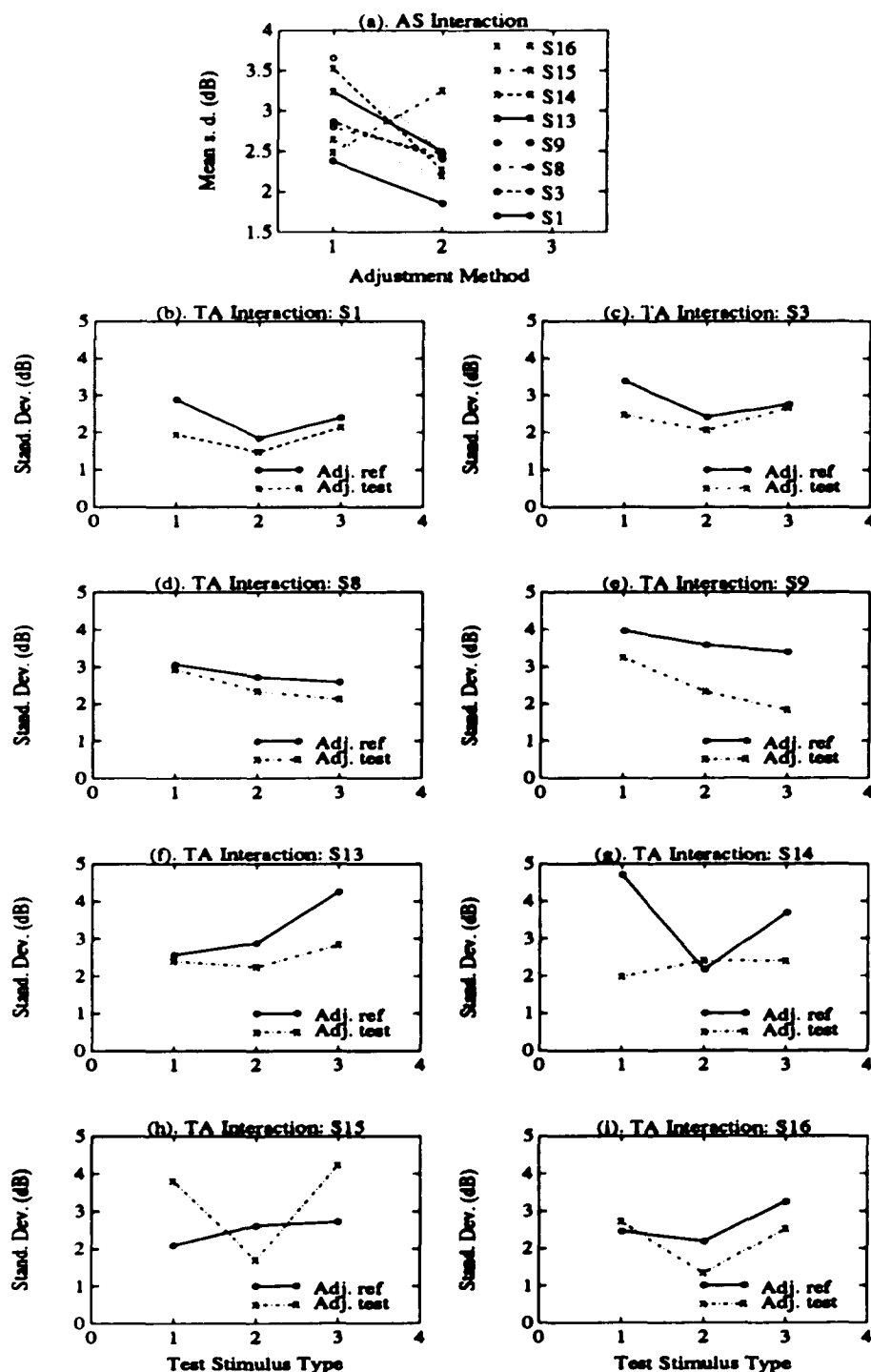


Figure 6.8: Effect of Interactions AS and TAS for Each Subject on Test-retest Variability. In Panel (a), the adjustment method 1 and 2 on abscissa represent the adjusting-reference-stimulus method and the adjusting-test-stimulus method, respectively. In Panels (b) to (i), the test stimulus type 1, 2, and 3 on abscissa represent SP, NBN150, and SQW, respectively. The ordinate represents the mean standard deviations averaged across variables L (stimulus level), C (experimental condition), and R (reference stimulus type).

variables, L, T, R, A, and C, although there were some variations with S and two interactions involving S. This result indicated that the homogeneity-of-variance assumption for an analysis of variance on the measured loudness levels could be made safely, provided caution is taken with respect to the significance of inter-subject differences and interactions of the form AS and TAS. A repeated measures design (over S) represents a reasonably safe approach for such an analysis.

6.4.2 Analysis of Experimental Variables

The analysis of test-retest variability in the previous section showed that the homogeneity-of-variance assumption for a repeated measures analysis of variance on the main experimental variables appeared quite reasonable. A repeated measures analysis of variance on the means of three replications for each test condition was thus performed on the six experimental variables, Stimulus Level (L), Test Stimulus Type (T), Reference Stimulus Type (R), Adjustment Method (A), Experimental Condition (C) and Subjects (S). These variables and their respective levels are summarized in Table 6.2. The results of the analysis are summarized in Table 6.4. Graphical plots for the mean response levels of each significant variable or interaction are provided in Appendix A.6.

The results showed that of the six experimental variables, L, C, and S, were statistically significant ($p \leq 0.001$ for each). Among the interactions common to all subjects, LC, AC, LTA, LTC, LAC, TAC, and LTAC were found to be statistically significant ($p \leq 0.001$). The main effects L, and C, and 2-way interactions LC and AC are shown graphically in Fig. 6.9. Graphical representations of the more complex significant interactions are shown in Appendix A.6.1. Of the interactions involving subjects, LS, TS, LTS, AS, LAS, TAS, LTAS, RS, RAS, CS, LCS, TCS, ACS, LACS, and TACS were found to be statistically

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

Analysis of Experimental Variables													
Grand Mean: 57.990608 dB SL1k													
Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F - Ratio	Signf. Level		Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F - Ratio	Signf. Level	
L	67123.75781	2	33561.87891	1023.009	0.001	•	LS	459.29828	14	32.80702	7.199	0.001	•
T	439.98755	2	219.99377	4.206	0.036		TS	732.29395	14	52.30671	11.479	0.001	•
LT	153.57730	4	38.39433	2.627	0.055		LTS	409.25909	28	14.61640	3.208	0.001	•
R	46.56518	1	46.56518	1.933	0.206		RS	168.59355	7	24.08479	5.285	0.001	•
LR	56.35675	2	28.17837	3.623	0.053		LRS	108.89089	14	7.77792	1.707	0.080	
TR	31.45090	2	15.72545	2.895	0.088		TRS	76.05645	14	5.43260	1.192	0.307	
LTR	59.90982	4	14.97746	2.505	0.064		LTRS	167.43492	28	5.97982	1.312	0.191	
A	4452.28613	1	4452.28613	2.160	0.183		AS	14431.91406	7	2061.70190	452.433	0.001	•
LA	1081.13281	2	540.56641	9.639	0.003		LAS	785.17310	14	56.08379	12.307	0.001	•
TA	3486.37817	2	1743.18909	11.199	0.002		TAS	2179.17871	14	155.65562	34.158	0.001	•
LTA	2083.58716	4	520.89679	18.193	0.001	•	LTAS	801.68219	28	28.63151	6.283	0.001	•
RA	628.42920	1	628.42920	6.287	0.039		RAS	699.70026	7	99.95718	21.935	0.001	•
LRA	55.34697	2	27.67349	4.682	0.027		LRAS	82.74184	14	5.91013	1.297	0.238	
TRA	8.96885	2	4.48443	0.939	0.583		TRAS	66.83980	14	4.77427	1.048	0.423	
LTRA	16.23335	4	4.05834	1.503	0.228		LTRAS	75.61906	28	2.70068	0.593	0.933	
C	45439.43359	2	22719.71680	135.422	0.001	•	CS	2348.77148	14	167.76939	36.816	0.001	•
LC	22267.19727	4	5566.79932	389.455	0.001	•	LCS	400.22659	28	14.29381	3.137	0.001	•
TC	439.24738	4	109.81184	3.796	0.014		TCS	809.93903	28	28.92639	6.348	0.001	•
LTC	436.84683	8	54.60585	7.073	0.001	•	LTCS	432.36749	56	7.72085	1.694	0.025	
RC	15.75820	2	7.87910	1.005	0.393		RCS	109.78489	14	7.84178	1.721	0.077	
LRC	21.09154	4	5.27289	0.847	0.509		LRCS	174.33762	28	6.22634	1.366	0.159	
TRC	41.12732	4	10.28183	2.862	0.041		TRCS	100.59767	28	3.59277	0.788	0.751	
LTRC	6.41166	8	0.80146	0.200	0.989		LTRCS	224.60640	56	4.01083	0.880	0.683	
AC	7859.99707	2	3929.99854	15.552	0.001	•	ACS	3537.71558	14	252.69397	55.453	0.001	•
LAC	1131.04785	4	282.76196	9.410	0.001	•	LACS	841.40771	28	30.05028	6.594	0.001	•
TAC	611.50586	4	152.87646	7.747	0.001	•	TACS	552.51025	28	19.73251	4.330	0.001	•
LTAC	566.23480	8	70.77935	9.203	0.001	•	LTACS	430.69693	56	7.69102	1.688	0.026	
RAC	22.13496	2	11.06748	0.804	0.529		RACS	192.69835	14	13.76417	3.020	0.002	
LRAC	44.74043	4	11.18511	2.158	0.099		LRACS	145.15619	28	5.18415	1.138	0.333	
TRAC	11.41476	4	2.85369	0.707	0.596		TRACS	113.02294	28	4.03653	0.886	0.629	
LTRAC	59.58456	8	7.44807	1.634	0.136		LTRACS	255.18750	56	4.55692			
S	4027.64624	7	575.37805	126.265	0.001	•	TOTAL	194639.12500	863				

Table 6.4: , L (Stimulus Level), T (Test Stimulus Type), R (Reference Stimulus Type), A (Adjustment Method), C (Experimental Condition), and S (Subject).

significant ($p \leq 0.001$). The main effect S and 2-way interactions, LS, TS, AS, RS, and CS are shown in Fig. 6.11. Graphical representations of more complex interactions are shown in Appendix A.6.2. Of all these significant effects, the largest ones were the main effect L, followed by interactions AS, LC, and the other two main effects C and S. All main effects and significant two-way interactions were subjected to a post-hoc analysis⁶ (Tukey, 1949).

Effects Common to All Subjects

An examination of the main effects L and C, and significant two-way interactions LC and AC in Fig. 6.9, and significant complex interactions common to all subjects shown in Figures A.2 through A.4 show that:

(1) The effects on the measured responses involving variables L, C, and A were as expected and consistent with the design of this experiment. The effects caused by the stimulus level (L) was obvious (see Panel (a) in Fig. 6.9), but its effect in conjunction with other variables would be of interest later. The effect caused by the experimental condition (C) (see Panel (b) in Fig. 6.9), was mainly due to differences of tasks between the conditions. The effect caused by the adjustment method (A) did not necessarily reflect the psychophysical behavior of the subjects. What variable A reflects depended on the experiment condition. In condition Q, the effect of A would reflect the psychophysical behavior of the subjects, because the task and the two stimuli being matched for loudness were *symmetrical*. By symmetrical, it is meant that both the test and the reference were in quiet, and the tasks in the two adjustment methods were the same, one stimulus was fixed in level while the other was adjusted to match in loudness. In conditions P and O,

⁶The post-hoc analysis of significant interactions did not extend beyond two-way interactions because the more complex interactions involved too many contrasts for a reliable analysis.

the effect of A reflected largely the *asymmetrical* design of the task and of the two stimuli been matched. By asymmetrical, it was meant that (a) the test stimulus was in noise and the reference stimulus was in quiet, and (b) the tasks in the two adjustment methods were not the same. In one method, the test stimulus in noise was fixed in level, while in the other method, the reference stimulus was fixed in level. In one method, the loudness of the reference in quiet was adjusted, while in the other method, the partial- or the overall loudness of the test stimulus was adjusted. For this reason, the levels used to set the fixed stimulus in conditions P and O were different for the two adjustment methods, as mentioned earlier. For the same reason, the measured responses obtained by the two methods would be different in conditions P and O. Hence in this ANOVA, the effect of variable A should not cause particular alarm at this stage except for the Q condition. However, the differences between the two methods for conditions P and O provided a common ground against which the individual subject's mean responses could be compared.

(2) The post-hoc analysis of the AC interaction indicated that the mean responses obtained using the two methods of adjustment differed significantly only for conditions P and O, but not significantly for condition Q (see Panel (d) in Fig. 6.9). An examination of interaction TAC in Fig. 6.10 indicated that test stimulus SQW is the main cause of the difference in subjects' responses. Another interesting point observed in Fig. 6.10 was that the relative positions of the three test stimuli for the two methods were very similar across all three experimental conditions. The mean responses were obtained in Panel (d) by adjusting the reference stimulus and in Panel (e) by adjusting the test stimulus. The patterns in Panel (d) are similar to each other and those in Panel (e) are similar to each other, indicating that the relation between the three test stimuli seemed genuine, i.e. the

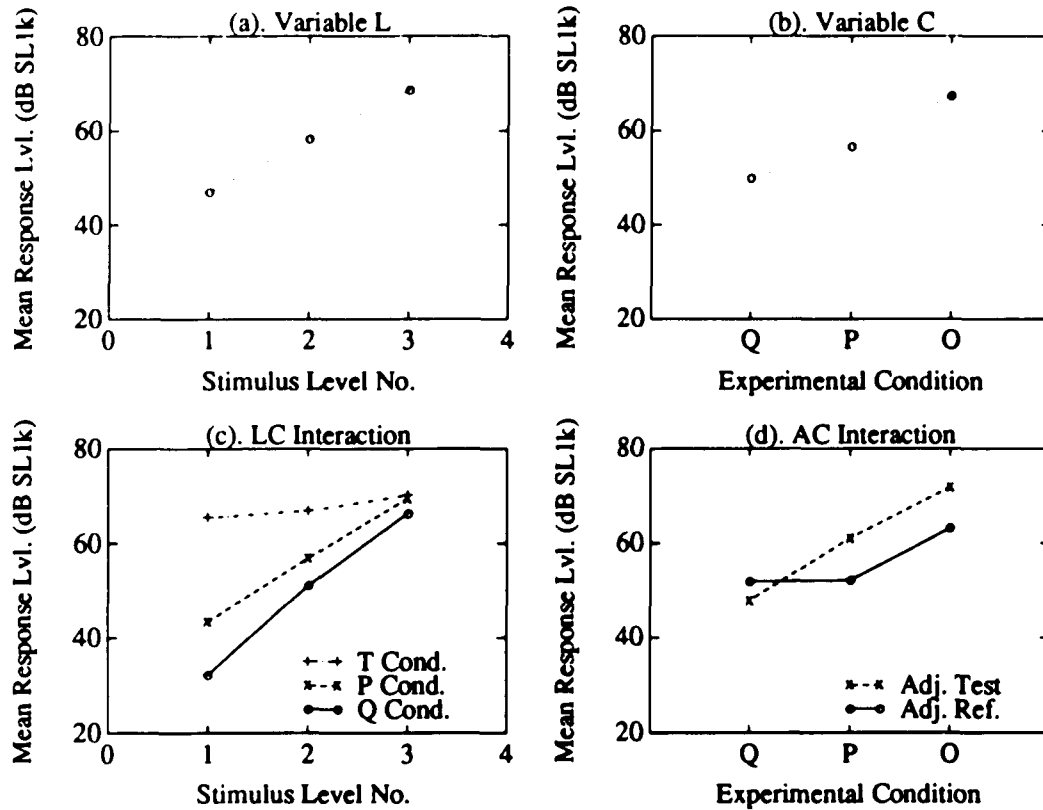


Figure 6.9: Effect of Stimulus Level, L, Experimental Condition, C, and Two-way Interactions, LC and AC. Panel (a) shows mean response levels for three stimulus levels. Each mean response is averaged over T, R, A, and C. Panel (b) shows mean response levels averaged over variables L, T, R, A, and S for each condition (C). Panel (c) shows mean response levels averaged over variables T, R, A, and S for each condition (C) at three stimulus levels (L). Panel (d) shows mean response levels averaged over L, T, R, and S for the two adjustment methods (A) in each condition (C).

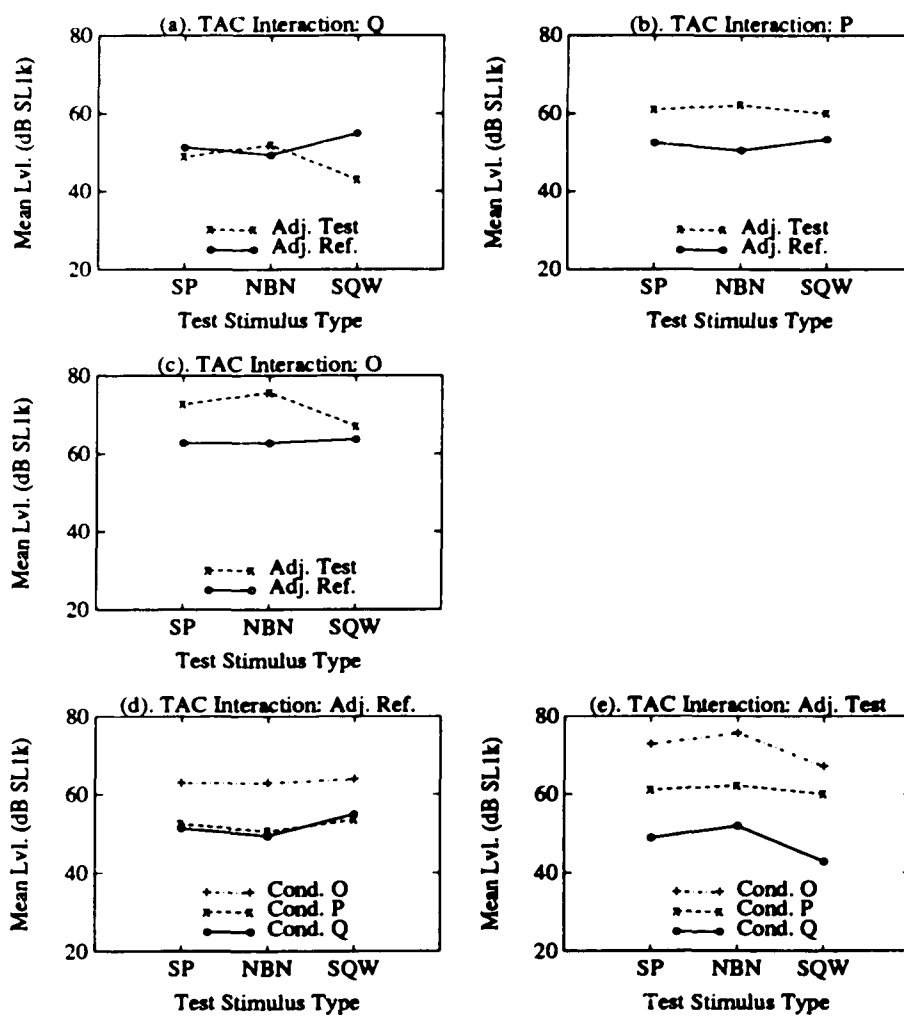


Figure 6.10: Graphical Representation of TAC Interaction. Mean response levels for interaction TAC shown in two ways. In Panels (a - c), mean response levels obtained using the two adjustment methods are plotted for three test stimuli in each panel for each condition: Panel (a) for condition Q, (b) for P, and (c) for O. In Panels (d - e), mean response levels for each test stimulus in each conditions are plotted with those obtained by adjusting the reference stimulus in Panel (d) and those obtained by adjusting the test stimulus in Panel (e).

differences in subjects' responses for different test stimuli are true psychophysical effects. For condition Q, Panel (d) shows that set at equal test stimulus level, SQW sounded slightly louder than SP and NBN150, and Panel (e) shows that SQW required less stimulus level to sound equally loud as SP and NBN150. In either case, SQW was demonstrated to behave differently from the other two test stimuli. This difference between the three test stimuli seemed to be affected by the background noise, i.e. the difference between the stimuli was reduced when the stimuli were partially masked, as in condition P. The difference between the test stimuli did not seem to be affected by the noise in (e). The tentative explanation would be that in this case, the test stimulus level was so much higher than the background noise that the noise was largely masked and the test stimulus dominated the overall loudness perception. Hence the individual characteristics of the test stimulus was not affected.

(3) Panel (c) of Fig. 6.9 shows that the mean responses to the three stimulus levels differ the most for condition Q, less for condition P, and very slightly for condition O. The results were affected by the background noise and the task performed. This interaction is as expected. An examination of the mean responses for the LTC interaction (see Fig. A.3) showed that the differences between the three test stimuli occurred at high stimulus level. For condition O (Panel (c)), all three levels were high, the three patterns showed the same order as in Panel (a) at high response level. The patterns for the three test stimuli were identical for condition P. This may be due to the reduced loudness of the test stimuli by the background noise.

(4) The graphical presentations of the mean responses for the interaction LTA in Panels (d) and (e) in Fig. A.4 shows how the pattern for SQW differed from those for SP and

NBN150 as a function of level.

Variations among Subjects

Another large main effect was the variation among subjects (S). This effect also included variations in test-retest variability among subjects, as reported in Section 6.4.1. Inter-subject differences in mean loudness levels were very large, however, (up to 6.3 dB) as compared to differences in test-retest variability on the order of 1 dB. Consequently, inter-subject differences were treated as if they consisted solely of differences in mean response levels. The post-hoc analysis at $p = 0.01$ significance level showed that subjects could be subdivided into five groups: [S9], [S13], [S16, S15], [S3, S14], [S14, S8], and [S8, S1], ranked in ascending order of their overall mean levels, which are shown in Panel (f) in Fig. 6.11. As noted earlier, subjects within a group did not differ significantly from each other in terms of their main loudness levels, but subjects in different groups did differ significantly. As can be seen from Fig. 6.11, most of the subjects differed significantly from each other. The lowest mean response was 54.1 dB, while the highest mean response was 60.4 dB.

Statistically significant interactions were also observed between S and each of the other variables (A, C, T, L, and R.) The interactions between S and each of these five variables are shown in Panels (a) to (e) of Fig. 6.11. In Panel (a), the abscissa represents the adjustment methods. Method 1 corresponds to the method of adjusting the reference stimulus, and method 2 to the method of adjusting the test stimulus. The patterns showed two groups of subjects. For one group (Subjects 1, 3, 13, 15, and 16), the mean responses obtained by adjusting the reference (1) were smaller than those obtained by adjusting the test (2), and vice versa for the other group (Subjects 8, 9, and 14). The magnitude of

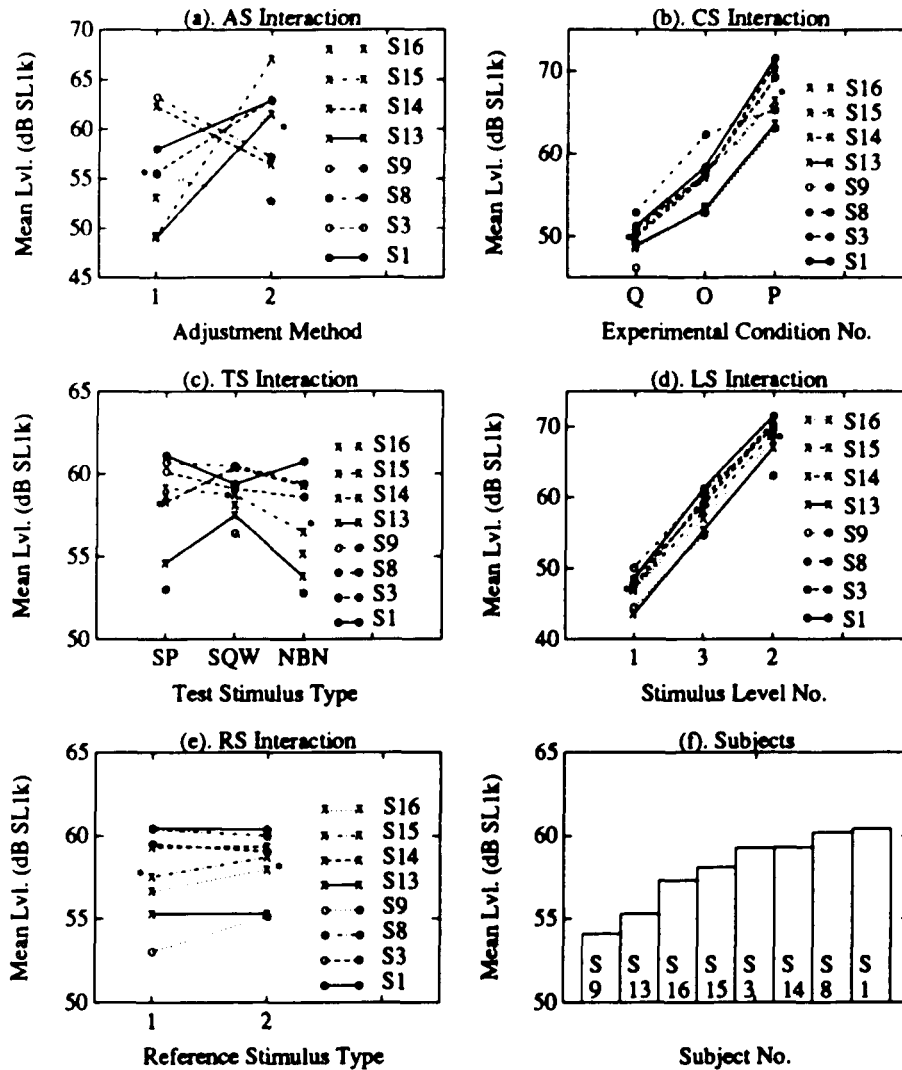


Figure 6.11: Interactions Involving the Variable S. Two-way interactions are shown separately in Panels (a) to (e) in descending order of the magnitude of the interaction as measured by F-ratio. In each panel, the abscissa shows the mean response averaged over the other four variables for each subject. In Panel (a), adjustment method 1 corresponds to adjusting the reference stimulus, and method 2 to adjusting the test stimulus. In Panel (b), conditions 1, 2, and 3 correspond to conditions Q, P, and O. In Panel (c), test stimulus types 1, 2, and 3 correspond to SP, NBN150, and SQW. In Panel (d), the stimulus levels 1, 2, and 3 correspond to the low-, mid-, and high-stimulus levels. In Panel (e), the reference stimulus types 1, and 2 correspond to PT and NBN120 respectively. The *'s represent the overall means across all subjects for each abscissa value. Panel (f) shows the mean response of each subject averaged across the five main variables, A, C, T, L, and R.

the difference between the two methods of adjustment varied from about 3 dB to 18 dB (Subject 15). In Panel (b), most subjects show similar patterns except S8, whose mean response for conditions Q and P (represented by conditions 1 and 2 of the abscissa) were the highest while that for condition O (condition 3) was smaller than the mean averaged over all subjects (represented by *). S9 had smallest mean responses uniformly across all three conditions. In Panel (c), test stimulus type 1, 2, and 3 correspond to SP, NBN150, and SQW. The overall spread of subjects' mean responses for NBN150 was smaller than those for SP and SQW. However, for SP, 6 out of 8 subjects seemed to form a group with a spread of the mean responses less than 4 dB, while subjects' mean responses for SQW scattered uniformly over a large range. S9 and S13 had smallest mean responses. In Panel (d), the three stimulus levels correspond to the low-, mid-, and high-stimulus level. S9 had smallest mean responses uniformly across all three levels, and S13 also had smallest mean response at the low-level, but was similar to other subjects at high-level. In Panel (e), the reference stimulus types 1 and 2 correspond to PT and NBN120. Five out of eight subjects showed the mean responses obtained using PT or NBN120 as a reference were nearly the same, while the rest three subjects had smaller mean responses using PT as a reference. However, the mean responses of the eight subjects spread over larger range when PT was used as the reference (about 8 dB) than when NBN120 was used (about 5 dB). S8 and S13 had the smallest mean responses again.

Overall, interactions involving variable A, S8, S9, and S14 had distinctly different interaction patterns from the other subjects, in that the mean responses obtained by adjusting the reference were higher than those obtained by adjusting the test stimulus. However, the differences between these three subjects and others occurred mainly for test stimuli

SP and SQW, and not for NBN150; for all three stimulus levels; and for all experimental conditions. The magnitude of differences between subjects was large and distinct when variable A was involved, otherwise the differences between subjects were small and not well defined.

The other statistically significant interactions involving subject included LCS, TCS, RAS, TAS, LTS, LAS, ACS, LTAS, LACS, and TACS. These more complex interactions are described and represented graphically in Appendix A.6.2. For more detail, loudness growth patterns averaged over the two reference stimuli for each test stimulus for each condition by the two adjustment methods are shown in Figures A.19 through A.22.

6.5 Discussion

6.5.1 Sources of Subject Variations

It is well known that measurements of loudness are inherently variable from subject to subject. The analysis of variance in the preceding section showed that inter-subject differences were highly significant, and that there were many more statistically significant interactions involving the variable S than interactions common to all subjects. Two kinds of variation among subjects are possible, one was that a subject might have significantly different overall means but similar response patterns, e.g. S13; and the other was that a subject might have different response patterns but similar overall means, or a combination of the two, e.g. S8. S9 had both different overall means and different patterns.

In general, within a subject, the variations seemed small, indicating the self-consistency in a subject's responses, as could be observed in Figures 6.2 through 6.7. In these figures, most of the loudness growth curves (solid lines) were approximately parallel to each other

indicating that most subjects had similar loudness growth rates for the same test stimulus. Also shown in the figures, although not clear, was that solid lines representing the loudness growth curves of a subject usually demonstrate the same characteristics. For example, a subject whose speech loudness growth curve was shallow would also have shallow growth curves for NBN150 and SQW. If a subject's speech loudness growth curve, using PT as the reference, was located above all other subjects' speech loudness growth curves, then the same would be observed for NBN150 and SQW using either PT or NBN120 as the reference. This observation was also true for the shapes of the loudness growth curves.

It was of interest to identify possible causes of the variations among subjects. Three possible factors might have contributed to the large variations among subjects. One possible source was that the differences between the detection thresholds of 1-kHz pure tone and other stimuli might be inconsistent across subjects. For example, the difference between the detection thresholds of PT and SP for a subject might be substantially larger than that for another subject. To explore this possibility, five of the original 12 subjects were tested for their thresholds for all stimuli used in the experiment⁷. (The measuring procedure and the results are shown in Appendix A.4.) The results confirmed that there were moderate variations between the threshold differences among subjects. These variations in threshold differences could account for some of the differences in the loudness patterns between subjects. Figures 6.12 through 6.13 provide a comparison between loudness patterns represented in SL1k and in estimated SL for those five subjects. Fig. 6.12 represents a subset of Fig. 6.2 in which levels are specified in SL1k. Fig. 6.13 represents the same set of data with stimuli levels adjusted to reflect the true sensation level of each subject for each

⁷These subjects were the only five available.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

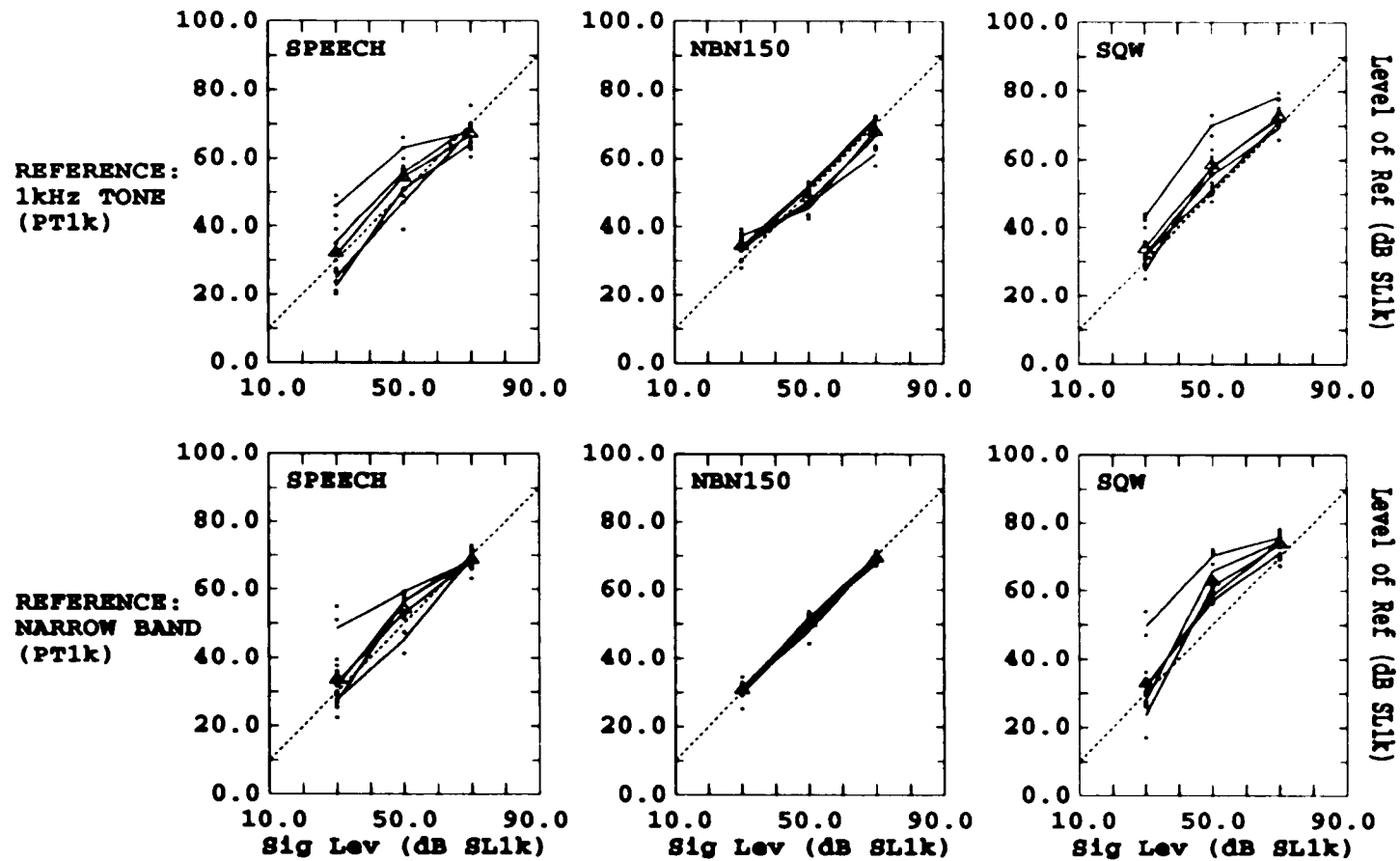


Figure 6.12: Original Loudness Growth Curves for Five Subjects. These curves are extracted from Fig. 6.2.

CHAPTER 6. LOUDNESS MATCHING EXPERIMENT

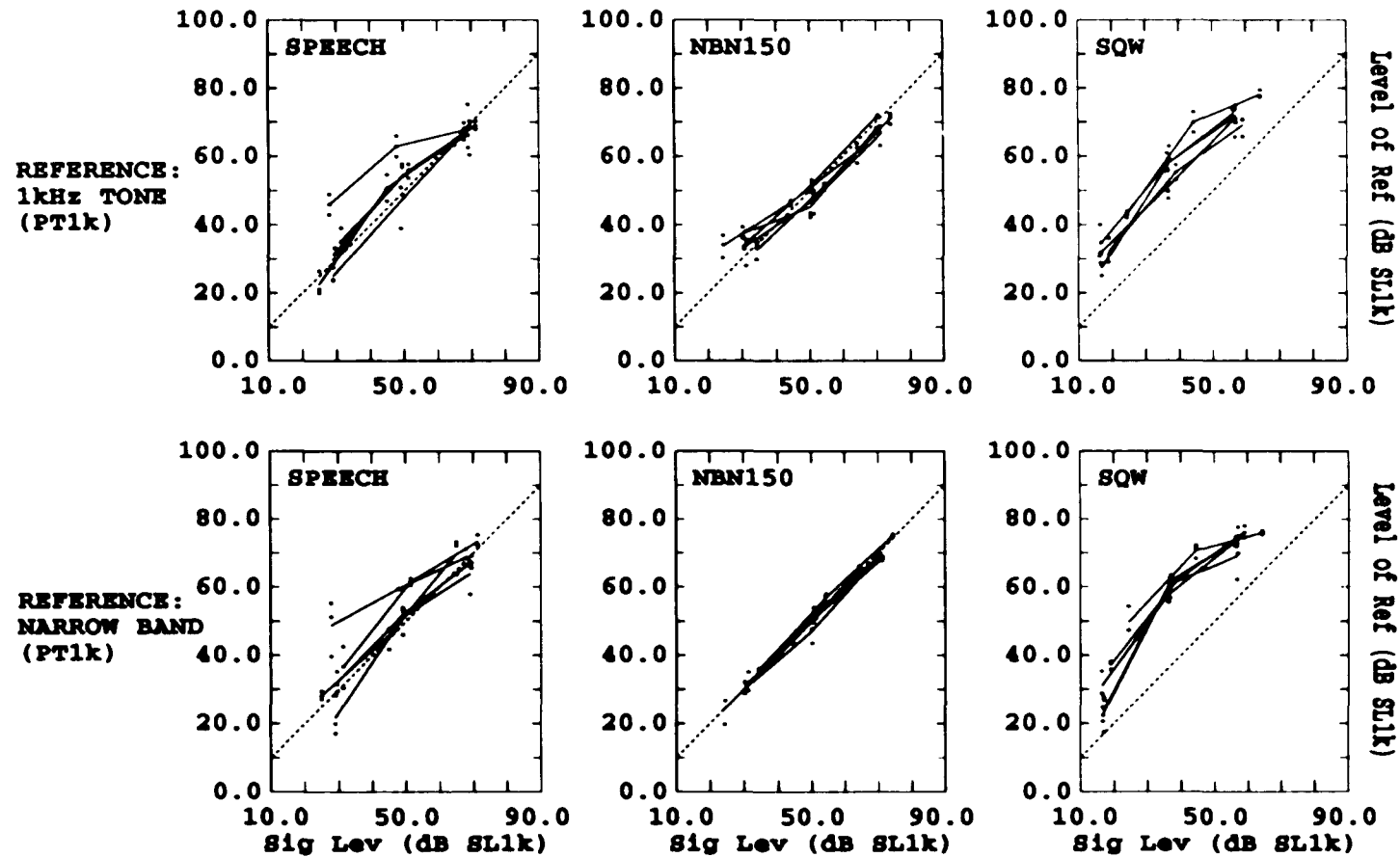


Figure 6.13: Loudness Growth Curves Corrected for the Individual Stimulus Thresholds for Each of the Five Subjects. The format is the same as Fig. 6.2.

stimulus. The comparison showed that the loudness growth curves showed a somewhat higher degree of similarity except for speech (SP), where one subject clearly differed from the other at mid- and low-level. Besides reducing inter-subject variability, the loudness growth curves based on SL also showed distinct differences between the square wave and the other two test stimuli.

The second possible source of variation was that different criteria might be used by different subjects to judge loudness when the test and reference stimuli sounded very different. Evidence in support of this interpretation was that the loudness growth curves were almost identical for all subjects when the two stimuli compared were virtually identical as in the case of NBN150 and NBN120. As the differences between two stimuli increased the subject variations increased correspondingly as in the case of NBN150 and PT. When the two stimuli were very different in spectrum shape and in temporal patterns (e.g. NBN120 and SP), the subjects showed the largest differences in their loudness growth functions. This interpretation was also evident from the figures corresponding to the significant interactions involving subjects in the second ANOVA. As discussed in Section 6.4.2, subjects S8, S9, and S14 behaved distinctly differently from other subjects, however not for NBN150, which was very similar to the reference stimuli.

Another possible explanation for the variation was that these subjects might differ from other subjects physiologically, in that their loudness summation mechanisms across channels differed from others. Therefore, stimuli requiring loudness summation, such as SP and SQW, would result in different loudness patterns from other subjects. In this case, other measurements, such as distortion products, might be useful.

6.5.2 Stimulus Level Adjustment

As demonstrated in the previous section, specification of levels in terms of their true sensation level provided more consistent patterns of loudness growth across subjects. Consequently, the loudness growth patterns were replotted using estimates of the true SL for each stimulus type. The averaged detection thresholds of the five available subjects for the six stimuli were used to estimate the difference between SL and SL_{1k} for each stimulus type and to adjust the loudness growth patterns accordingly. The method and the results of threshold measurement are described in detail in Appendix A.4. Table 6.5 shows the differences of the averaged thresholds of the six stimuli from that of PT. A positive value

Threshold difference from 1-kHz PT in dB					
Speech	NBN150	SQW	PT	NBN120	MPN
1.1	-0.6	10.7	0.0	-0.5	4.1

Table 6.5: The signal thresholds are obtained from 5 subjects with one subject measured twice.

means that the stimulus has higher threshold than PT, and vice versa for a negative value. Table 6.5 showed that speech, SQW, and modified pink noise (MPN) had higher thresholds than PT, and the narrow-band noise (NBN) had lower thresholds. The 10.7-dB difference between SQW and PT indicated that the SQW was presented to the subject 10.7 dB below the targeted level in the experiment. Therefore, to represent the stimulus level in true SL, 10.7 dB had to be subtracted from SQW's levels in SL_{1k}⁸. These changes are incorporated in the revised plots of the averaged loudness levels shown in the next section.

6.5.3 Averaged Loudness Levels

⁸That is, the stimulus levels was corrected to show dB re the averaged detection threshold of the stimulus and not of 1-kHz pure tone.

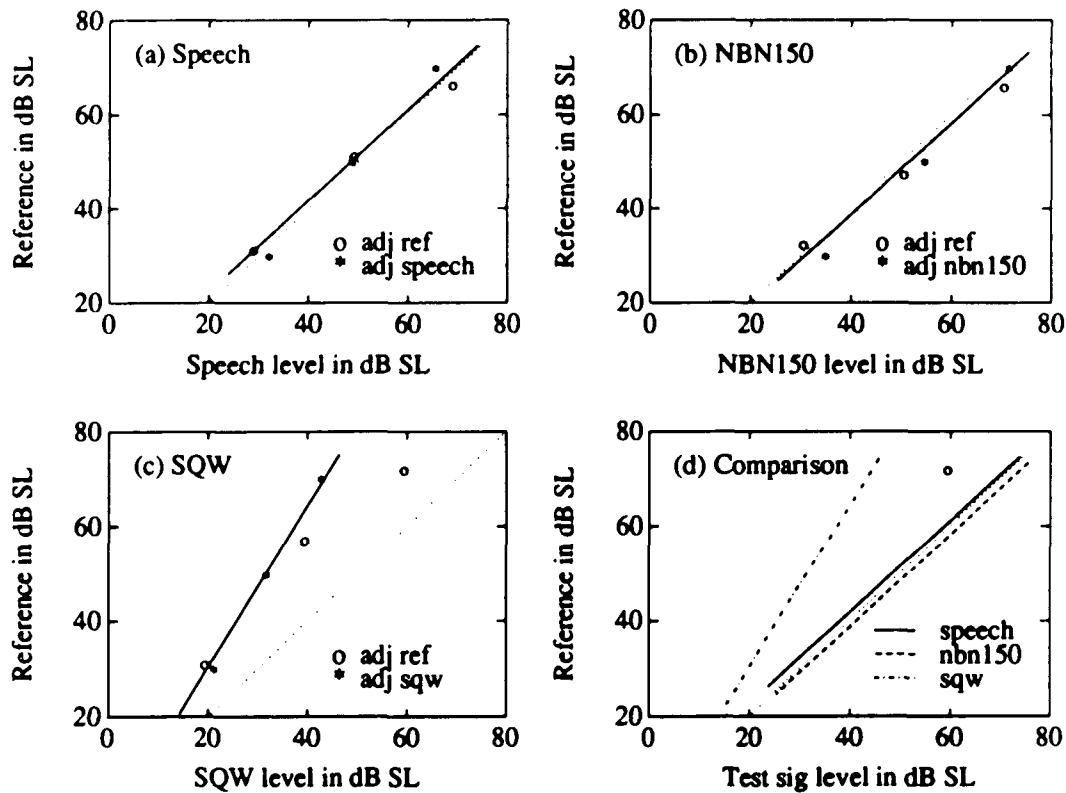


Figure 6.14: Averaged Loudness Levels in Quiet; Stimulus Levels Specified in SL. Data averaged over Subjects and Reference Stimulus (PT and NBN120) are shown. The open circles, o, represent the averaged loudness levels measured by adjusting the reference stimulus. The asterisks, *, represent those measured by adjusting the test stimulus. The solid straight lines fitted to the data, using linear regression, represent the estimated loudness-growth functions. Panels (a - c) correspond to the three test stimuli (SP, NBN150, and SQW) respectively. Panel (d) compares the loudness growth functions of the 3 test signals. Note that the straight lines in panels (a) and (b) are fitted to both sets of data, o and *, while the straight line in panel (c) is fitted to the lowest five loudness data points, two o's and three *'s. The open circle at high-level is left alone.

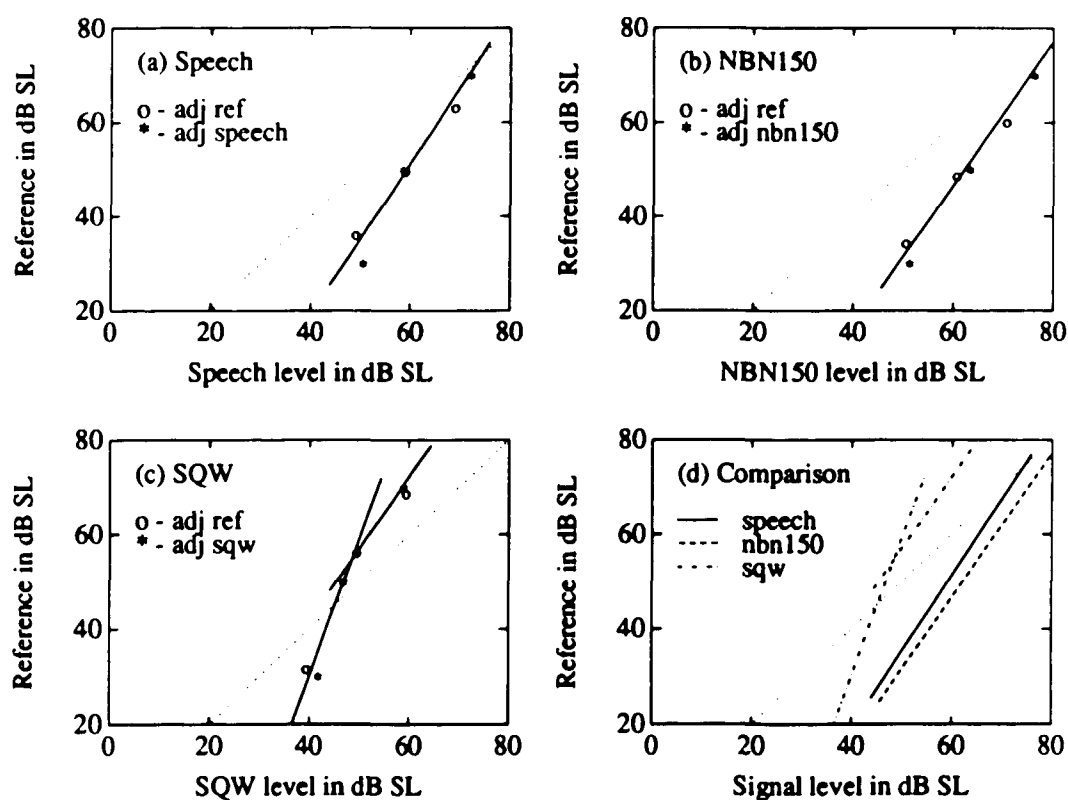


Figure 6.15: Partial-loudness Levels for Test Stimuli in Noise; Stimulus Levels Specified in SL. The loudness of the test stimulus in noise was matched to the reference stimulus. Nomenclature and labelling is identical to that for Fig. 6.14.

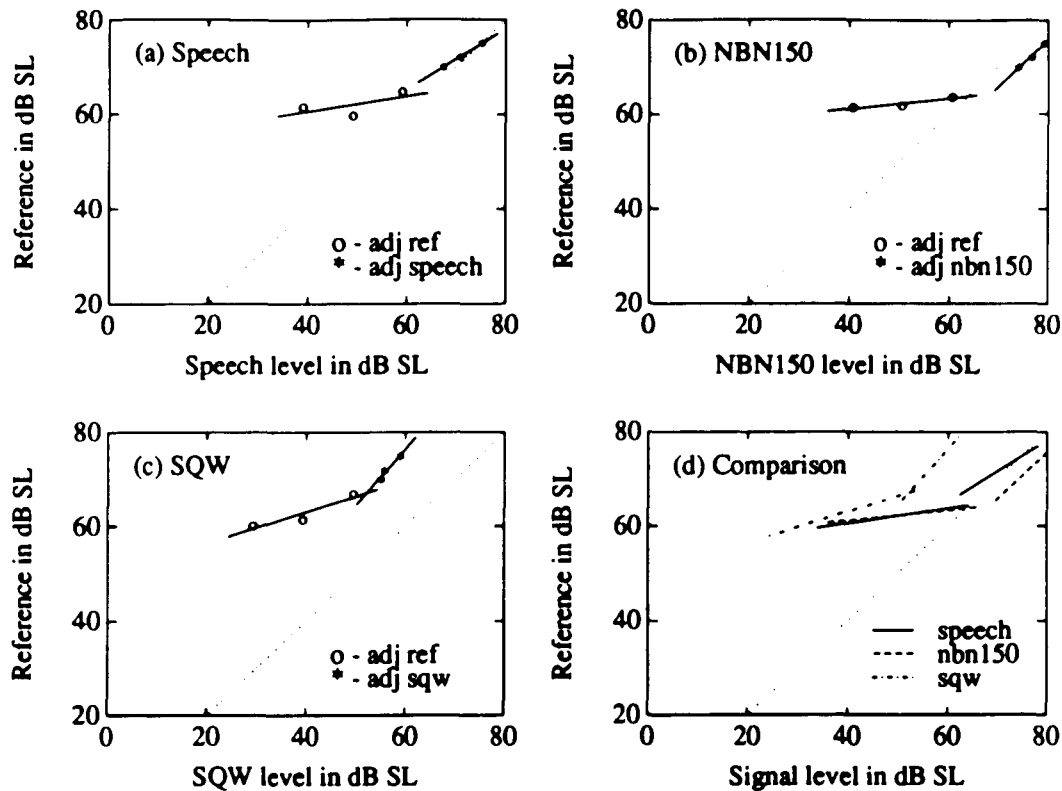


Figure 6.16: Overall Loudness of Test Stimuli in Noise; Stimulus Levels Specified in SL. The loudness of the test stimulus plus noise was matched to the reference stimulus. Nomenclature and labelling is identical to that for Fig. 6.14.

Figures 6.14 through 6.16 depict the averaged loudness levels as a function of stimulus level (L), test stimulus type (T), adjustment method (A), and experimental condition (C). The three figures correspond to the three conditions; Fig. 6.14 for condition Q (in quiet), Fig. 6.15 for condition P (Partial-loudness), and Fig. 6.16 for condition O (Overall-loudness). In each figure, Panels (a) to (c) corresponds to the three test stimuli; Panel (a) for speech (SP), Panel (b) for narrow-band noise (NBN150), and Panel (c) for square wave (SQW). Panel (d) compares the derived loudness growth patterns for the three test stimuli. The derivation of the loudness growth patterns will be discussed in the next section.

In each panel of the figures, the abscissa represents the test stimulus level and the ordinate the reference stimulus level. Both levels are shown in dB estimated sensation level. The two symbols in Panels (a) to (c) represent the averaged loudness levels obtained by two adjustment methods. The open circles, $\circ$, represents the loudness levels obtained by adjusting the reference stimulus, and the asterisks, $*$, represents those obtained by adjusting the test stimulus. The dotted diagonal lines function as a guide line. Any data point falling on the dotted diagonal line indicates that the test stimulus and the reference stimulus sound equally loud at the same SL. A data point located above the dotted diagonal line indicates that the test stimulus sounds louder than the reference at the same SL; and vice versa for a data point located below the dotted diagonal line.

Each averaged loudness level represented by a symbol was obtained by taking the arithmetical mean of the measured loudness levels for the two reference stimuli, PT and NBN120, and averaged over subjects. Thus, the ordinate in each panel of the figures represents the averaged reference level over PT and NBN120. The number of subjects used to derive the averaged loudness levels varies with condition. As indicated in Section 6.2.1, one subject had missing data for the P condition, and 3 subjects had missing data for the O condition. Data from 8 subjects, who completed all measurements, were used in the main data analysis. A separate analysis on the incomplete data of the remaining subjects was also performed. The same basic results were obtained from these subjects, and therefore, their data were included in calculating the overall averaged loudness levels. That is, data for the Q and O conditions from S6 were included, and data for Q and P conditions from S4, S5, and S7 were included. Their data had similar standard deviation to the mean standard deviation and their mean loudness levels fell within the range of the means of

other subjects.

6.5.4 Regression Effect

It has been observed (Stevens, 1961, Scharf, 1959, and others) and studied (Stevens and Greenbaum, 1966) that in an equal-loudness matching task, the adjusted levels tended to gravitate toward the mid-level. This tendency has been attributed to the subjects' preference for more comfortable levels. That is, in a loudness matching task, a stimulus was fixed in level (standard stimulus) and the other stimulus was adjusted in level (variable stimulus). If the standard stimulus was at a very low level, the subjects tended to assign a higher loudness to the standard, and hence to adjust the variable stimulus to a higher level. On the contrary, if the standard was at a very high level, the subjects tended to assign a lower loudness to the standard stimulus, and hence to adjust the variable stimulus to a lower level. The subjects tended to *regress* the adjusted stimulus toward more comfortable level. This tendency is termed *regression effect*.

This regression effect is observed in the data from this experiment. In Panels (a - c) of Fig. 6.14 and Fig. 6.15, imagine a line connecting the three o's, call it line A, and imagine another line connecting the three *'s, call it line B. The slope of line A is shallower than that of line B in all cases. The two lines intersect somewhere in the mid-level. The shallower slope of line A indicates that when the reference stimulus is adjusted, the adjusted reference levels tend to be drawn toward the mid-level. The steeper slope of line B indicates that when the test stimulus is adjusted, the adjusted test stimulus levels also tend to be drawn toward its mid-level. This observation demonstrates the classical regression effect.

In order to take the regression effect into account, it is advisable to not only specify the averaged loudness level (averaged over both methods of adjustment) but also the magnitude

of the difference in measured loudness levels. In Fig. 6.14 and Fig. 6.15, the loudness growth patterns (the solid lines) were derived using a linear regression fit to the averaged loudness levels from both adjustment methods (o's and *'s), except for square wave in panel (c). In Fig. 6.14, a straight line for square wave was fitted to the lowest five loudness data points and in Fig. 6.15 two line segments were used to fit the data. The individual fittings are discussed in the following section, Sec. 6.5.5.

The magnitude of the regression effect was on the order of 2.1 dB for Q condition and 2 dB for P condition. The minimum and the maximum deviations from the least-square fitted lines were 0 and 6.3 dB for Q condition, and 0 and 3.6 dB for P condition. The magnitude of the regression effect also varied with subjects. Since the adjustments for the true sensation level and the least-square fit were not applied to individual subject's loudness levels, the exact magnitude of the regression effect for each subject was not available. However, Figures A.19 through A.22 showed the averaged loudness patterns of each test stimulus measured using the two adjustment methods in each experimental condition for each of the 8 subject with complete data set. A comparison across the subjects for each test stimulus for each condition should provide a sense of variations between the magnitudes of regression effect.

6.5.5 Loudness Growth Curves

In-quiet Condition (Q)

Fig. 6.14 Panels (a) and (b) show average loudness level data. Straight lines have been fitted to these data, as described in the preceding section, to obtain the estimated loudness-growth curves. Curve SP is slightly to the left of the diagonal line with a slope of 0.966.

and curve NBN150 is slightly to the right with a slope of 0.972. Both curves are very similar to the dotted diagonal line, a slope of 1. However a 2-way ANOVA with five test stimulus level for the two loudness-growth curves SP and NBN150 and for the diagonal line was performed to examine the significance of the variations between the two curves and the diagonal line. The results show that the differences between the two curves and the diagonal line are statistically significant ($p \leq 0.001$). The post-hoc analysis shows that each curve is significantly different from the other two. If the data are collapsed over the five levels for each curve, the results show the differences between the curves are not significant.

In Panel (c), the averaged loudness data (o's and *'s) shows that the loudness-growth curve for SQW is very different from those for SP and NBN150. All data points locate to the left of the dotted diagonal line very substantial distances, and the loudness data obtained by two adjustment methods are close to each other at low level and are far apart at high stimulus level. It is apparent that the loudness growth rate of SQW is not proportional to that of the reference stimulus above approximately 45 dB SL of SQW. Furthermore, there is not enough data to indicate how the loudness growth of SQW behaves above 45 dB SL. A straight line does not seem to be appropriate for fitting the entire stimulus range of SQW. Instead, a straight line is fitted to the lowest five loudness data points, i.e. the two open circles at low-, and mid-level, and the three asterisks, and the open circle at high-level is left alone. The slope of the straight line is 1.7. The open circle at the high-level seems to indicate a drastic reduction of loudness growth rate for SQW relative to the reference stimuli.

Panel (d) shows the loudness growth curves plotted together for comparison. The

comparison shows that (1) the loudness growth curve for SQW is much different from those for SP and NBN150; (2) up to 45 dB, the loudness of SQW grows much faster than SP and NBN150; (3) SP has similar loudness growth rate as NBN150 but is judged to be slightly louder than NBN150 at the same SL. Based upon the ANOVA mentioned earlier, the difference between the loudness-growth curves of SP and NBN150 is significant. However, based on the s.d.'s of the mean thresholds for SP and NBN150, which are 5.3 dB and 6.3 dB, respectively, the difference between SP and NBN150 for the same SL, which is 3.4 dB across the intensity range, does not seem significant. An examination of the thresholds of the five subjects for SP and NBN150 shows that four out of five have lower NBN150 detection thresholds.

Partial-loudness Condition (P)

Fig. 6.15 is identical to Fig. 6.14 except the loudness levels are measured for the test stimuli embedded in the modified pink noise (MPN) which is set at 50 dB SL_{1k} or 46 dB SL (as estimated in Appendix A.4). The solid straight lines have been fitted by the method of linear regression to the averaged loudness data. Loudness judgments were of the test stimuli partially masked by noise; i.e. partial-loudness levels were obtained. Panels (a - c) show the estimated partial-loudness growth curves for speech (SP), narrow-band noise (NBN150), and square wave (SQW) in MPN respectively, and panel (d) compares the three partial-loudness growth curves. These curves show the effect of partial masking by MPN on loudness growth. The partial masking increases the detection threshold of the stimuli, reduces the dynamic range of hearing, and increases the loudness growth rates of the stimuli.

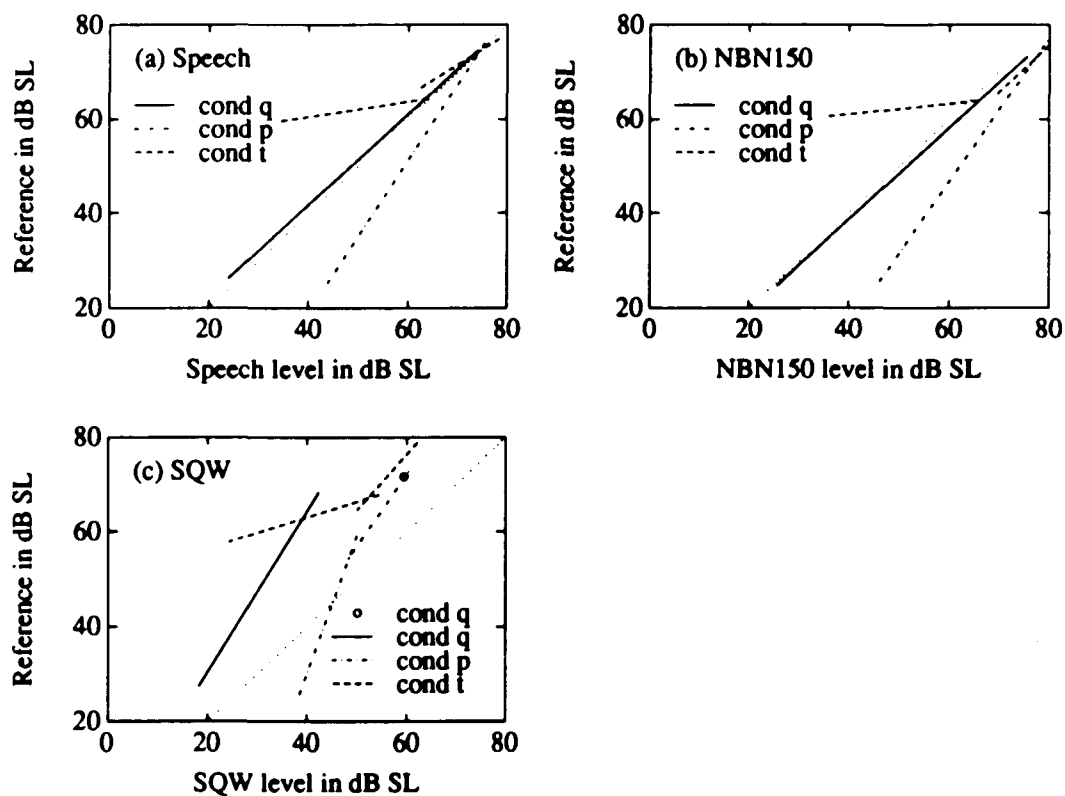


Figure 6.17: Comparisons of Loudness Growth Curves for Speech (SP), Narrow-band Noise (NBN150), and Square Wave (SQW) in Three Test Conditions. The three test conditions are Quiet (Q), Partial-loudness (P), and Overall-loudness (O). The curves for the three test conditions are represented by solid lines and a open circle, by dash-dotted lines, and by dash lines respectively.

Comparing these partial-loudness growth curves to the corresponding loudness curves in quiet, as shown in Fig. 6.17, shows that (1) the detection thresholds of all three test stimuli have increased by similar amount; (2) the shape of these curves remains the same; (3) the slopes of these curves are steeper. The loudness growth curves for SP and NBN150 are approximated quite well by straight lines (over the stimulus range considered in this experiment), indicating that the loudness growth of SP or NBN150 is linearly proportional to that of the reference stimuli. The loudness growth curve for SQW is approximated by two straight lines segments. The lower segment is fitted to the 4 data points at the lower sensation levels and the upper segment is fitted to the 4 data points at the upper sensation levels.

The effect of the modified pink noise (MPN) on the perception of loudness is large near the masked thresholds of the test stimuli and diminishes rapidly with increasing sensation level. The loudness of SP and NBN150 approach the loudness levels in quiet at about 75 dB SL as shown in Panels (a) and (b) of Fig. 6.17. The loudness of SQW grows extremely rapidly with stimulus intensity from 40 to 50 dB SL, and slows down from 50 to 60 dB SL. The upper segment show a reduction of loudness growth rate similar to that of the SQW in quiet, as indicated by the open circle for Q condition. There are not enough data points to illustrate the exact behaviour of the loudness growth from 50 dB SL and up. Nonetheless, it is clear that SQW has a much faster loudness growth rate at low sensation levels than the reference stimuli.

The comparison of the three partial-loudness curves in Panel (d) of Fig. 6.15 shows a similar relationship between three stimuli as in the Quiet condition. The relative locations of the three loudness-growth functions is the same for both the Quiet and Partial-loudness

conditions. The loudness growth curve for SP lies slightly above that for NBN150, whereas the curve for SQW is located well above either of these two curves. This indicates that at the same SL, SP was judged to be louder than NBN150, and SQW was judged to be much louder than both SP and NBN150. The rate of loudness growth of NBN150 is similar but slightly slower than that of the speech as shown by the lightly greater slope of the NBN150 curve. (The slopes of the SP and NBN150 curves are 1.6 and 1.5 respectively.) The upper segment of SQW, with a slope of 1.53, is parallel to the loudness curves of SP and NBN150 indicating that the loudness growth rate of SQW in the range of 50 to 60 SL is the same as that for SP and NBN150. The slope of the lower segment of SQW is 2.92. The effect of the MPN is similar across all three test stimuli.

To examine the significance of the difference between SP and NBN150, a 2-way ANOVA is performed on the two loudness-growth curves. The result of the analysis indicates the difference between SP and NBN150 is statistically significant.

Overall-loudness Condition

Fig. 6.16 depicts the overall-loudness levels and loudness growth curves of the test stimuli plus the background noise (MPN), i.e., data are for judgments of overall loudness of test stimulus plus noise. As before, the level of the MPN was 46 dB SL. The nomenclature and the layout of this figure are the same as in Fig. 6.14 and Fig. 6.15 except the regression effect was found to be so large that separate curves are fitted to data obtained by the two methods of adjustment. As before, the open circles indicate that the reference stimulus was adjusted and the asterisks indicate that the test stimulus was adjusted. The reason for the non-overlapping of the two sets of data was twofold. One reason was that the

overall-loudness growth curve of the test stimulus with MPN was not apparent, hence it was difficult to determine which stimulus levels to use while still remaining within the constraints of not going so high as to irritate subjects' hearing, and not too low so as to not cover the intensity range of interest. Another reason was that the first three subjects found it impossible to match the loudness when the reference stimulus was set at 65 dB SL_{1k}, as described in Section 6.2.1.

The curves fitted to the open circles have extremely shallow slopes, 0.16, 0.11, and 0.33 for SP, NBN150, and SQW, respectively. As the stimulus levels of SP, NBN150, and SQW decrease, the loudness curves seem to plateau at about 60 dB loudness level (reference level at 60 dB SL), which is approaching the loudness level of MPN alone. An increase of 20 dB for the test stimulus levels gives an increase of only 5 dB in the overall-loudness level. Note that the test stimuli embedded in the MPN are audible at the levels presented, and the loudness increases with level are also audible. The extremely slow change of the overall loudness seems to indicate a trading between the loudness of the test stimulus and the MPN due to mutual masking of the two stimuli. That is, at the low SL, most of the test stimulus energy is masked out by the MPN, so the MPN is the major contributor to the overall loudness. But as the test stimulus level increases, the increasing test stimulus energy starts to mask out the MPN, therefore, cancelling out the contribution of the increasing test stimulus loudness toward the increase of the overall loudness. Therefore, the overall loudness changes only by a small amount, if at all.

The upper line segments, fitted to the asterisks, have much steeper slopes, 0.65, 0.98, and 1.18 for SP, NBN150, and SQW, respectively, than the lower line segments. There seems to be a quick transition point between the two line segments for each test stimulus.

At the transition point, the overall-loudness growth rate changes dramatically. The exact pattern around the transition point requires further study. Above the transition points, the fitted lines show that the overall-loudness growth rates increase rapidly toward those of the corresponding test stimuli in quiet condition (Q) as depicted in Fig. 6.17. The merging of the corresponding loudness curves for conditions Q and O seems to indicate that the test stimuli are the main contributors to the overall loudness at high SL, 60 and 65 dB SL for SP and NBN150 and 50 dB SL for SQW. The subjects' comments confirm that within this small range of 5 dB, the test stimuli became so dominant that the overall-loudness matching was mainly based upon the loudness of the test stimuli only.

Chapter 7

Model Predictions II

In this chapter, predictions from ten versions of the EIH-based loudness model and from Zwicker's method of loudness calculation are compared with the loudness data obtained in the loudness matching experiment described in Chapter 6. In the first section of the chapter, three forms of loudness prediction of the EIH-LM models for the stimuli used in the loudness matching experiment are presented. The three forms are detection thresholds, loudness functions, and loudness level functions. The stimuli predicted include speech (SP), narrow-band noise with 150 Hz bandwidth (NBN150), square wave (SQW), 1-kHz pure tone (PT), narrow band noise with 120 Hz bandwidth (NBN120), and modified pink noise (MPN). For the loudness level data observed in the loudness matching experiment, data from two conditions, quiet (Q) and overall-loudness (O), are predicted by the EIH-LM models.

The ten versions of the EIH-based loudness model are EIH-LM3 through EIH-LM12. The parameters of the EIH-LM model that have been examined by the loudness predictions of these ten versions includes the definition of the overall loudness of a stimulus; the sensitivity threshold range of the five level detectors in each channel; the third integration

rule; the density of level detectors; and the manner of level detector distribution. See Chapter 4.3 for more details.

In the second section of the chapter, loudness level predictions based on Zwicker's method of loudness calculation are compared for SP, NBN150, and SQW in the Quiet and Overall-loudness conditions of the loudness matching experiment. The predictions were obtained using the computer programs developed by Zwicker, et al. (1984, 1985a) based on ISO 532 B (1975), an international standards for calculating loudness of stationary sounds. (see Chapter 3.1.3 for more details on ISO 532 B.) Loudness level functions for SP, NBN150, and SQW partially masked by MPN as in the Partial-loudness Condition were predicted using Zwicker's formulae (1987) developed to compute the loudness of the partially masked stimulus.

7.1 Predictions by EIH-Based Loudness Models

7.1.1 Predicted Detection Thresholds

The predicted detection thresholds for the six stimuli of the loudness matching experiment using the EIH-LM4 model are shown in column 2 of Table 7.1. Also shown are the means (column 3) and the standard errors of the means (column 4) for the detection thresholds measured for five subjects with one subject measured twice. The predicted threshold is defined to be the lowest sound pressure level for which the predicted loudness is greater than or equal to 20 model loudness unit (mlu) (see Section 4.2 for details). The difference between the observed and predicted thresholds for the respective stimulus are shown in column 5. The predicted detection thresholds lie below the observed mean thresholds by about 7 dB in average.

Detection Thresholds (dB SPL)							
Signal Type	Predicted	Observed		Obsv. - Pred.	Rel. Th. Diff. from PT		
		Mean	Std. E.		Pred.	Obsv.	Obsv.-Pred.
SP	12.0	16.1	2.2	4.1	3.0	1.1	-1.9
NBN150	5.0	14.4	2.6	9.4	-4.0	-0.6	3.4
SQW	18.0	25.7	2.4	7.7	9.0	10.7	1.7
PT	9.0	15.0	1.8	6.0	0.0	0.0	0.0
NBN120	4.0	14.5	2.5	10.5	-5.0	-0.5	4.5
MPN	14.0	19.1	2.0	5.1	5.0	4.1	-0.9

Table 7.1: Data are shown for six stimuli used in the loudness matching experiment. Predictions are for the EIH-LM4 model. The predicted detection threshold is defined to be the dB SPL at which the number of excitations at the lowest threshold detection is equal to 20, the maximum number of excitations for the lowest level detector.

An examination of these threshold data shows that of the most interest is the relative threshold differences between different stimuli rather than the absolute differences between the predicted and the observed threshold for the respective stimuli. The reason is that the predicted thresholds as a whole can be moved up or down relatively easier than the relative differences between the thresholds of different stimuli can be manipulated. Let the threshold of 1-kHz pure tone (PT) be the reference. By subtracting the predicted PT threshold from other predicted thresholds, the relative predicted threshold differences are obtained as shown in column 6. By subtracting the observed PT threshold from other observed thresholds, the relative observed threshold differences are obtained as shown in column 7. An examination of the relative differences of the observed detection thresholds in column 7 shows that NBN150, NBN120, and SP have similar thresholds to that of PT. The observed threshold of modified pink noise (MPN) is 4.1 dB higher, and that of square wave (SQW) is 10.7 dB higher. An examination of the relative differences of the predicted thresholds in column 6 shows similar patterns for SP, SQW, PT, and MPN. Each entry in column 8 is the difference between the corresponding entries in column 7 and column 6

(the observed and the predicted relative threshold differences). The differences in column 8 indicate a good match between the observed and the predicted thresholds in terms of the relative differences between stimuli, except for the two narrow bands of noise (3.4 and 4.5 dB). However, both 3.4 and 4.5 dB are within the confidence intervals¹ of the respective observed thresholds. The importance of the similarity in the relative threshold difference is in providing a possible explanation of how the loudness of complex stimuli are summed at threshold.

As mentioned in Chapter 4, the pure-tone thresholds are influenced by two parameters, the shape of the overall envelope of the cochlear filters and the sensitivity threshold of the lowest level detector in each channel. Manipulations of either parameter can shift the threshold levels up or down. Hence, the predicted thresholds being below the observed mean thresholds by an average of 7.1 dB is not a serious concern. Also mentioned in Chapter 4 is that the shape of the overall envelope of the cochlear filters is largely determined by the middle ear transfer function. Therefore, if the cochlear filters used in the model can be modified to reflect human middle ear transfer function (see Section 8.5.1 for more details), then the predicted thresholds can be improved. The effect of the sensitivity thresholds of the lowest level detectors in all channels on the predicted thresholds is demonstrated by EIH-LM5's predictions of 1-kHz pure tone and modified pink noise. The detection thresholds of the pure tone and modified pink noise predicted by EIH-LM5 are 11 and 17 dB SPL, an increase of 2 and 3 dB from those by EIH-LM4, respectively. The difference of EIH-LM5 from EIH-LM4 is in the threshold range of the level detectors and hence the sensitivity

¹The 95 percent confidence interval is 14.4 ± 6.7 for NBN150 and is 14.5 ± 6.4 for NBN120. Since the sampling size is small, 6 observations for each stimulus, the distribution of t with 5 degrees of freedom (Ferguson, 1981) is used to calculate the confidence intervals. The formula is $Mean \pm 2.57 \cdot Std.E.$, where 2.57 is the limit for 95 percent confidence interval.

thresholds of the lowest level detector. In EIH-LM5, the threshold range is changed to be a fixed value of 60 dB for all channel instead of reducing range with increasing characteristic frequency (CF) of the channels as in EIH-LM4. Thus, in EIH-LM5, the threshold levels of the lowest level detectors became 12 dB above the CF threshold of the respective filter for all channels instead of decreasing with increasing CF. This change reduces the differences between the predicted and the observed thresholds. In EIH-LM6, the sensitivity thresholds of the lowest level detectors are reduced slightly from those in EIH-LM5, and so are the predicted detection thresholds of the corresponding stimuli.

Both the observed and predicted detection thresholds for modified pink noise are 4 to 5 dB higher than those for 1-kHz pure tone respectively. A comparison of the power spectra of the two stimuli at equal overall sound pressure level shows that the energy of the 1-kHz tone is concentrated in the critical band centered at 1 kHz, while the energy of MPN is distributed approximately equally across critical bands up to 7.2 kHz, a property of the modified pink noise. As a consequence, the summed energy in the critical band centered at 1 kHz for the 1-kHz tone is higher than the summed energy in any critical band for modified pink noise, i.e. the critical band level for PT is higher than any critical band level for MPN. Hence, when measuring the detection thresholds, the overall SPL of MPN has to be higher for MPN to be detectable than for 1-kHz pure tone to be detectable. Thus, the observed detection threshold of modified pink noise would be higher than that of the tone or those of the two narrow bands of noises (NBN120 and NBN150) as well. In the EIH-LM models, the channel most sensitive to 1-kHz pure tone (cochlear filter with CF at 1 kHz) is like the critical band centered at 1 kHz. In order to excite the lowest level detector in this channel, the overall SPL of MPN has to be higher than that of the 1-kHz pure tone.

Thus the predicted detection threshold for MPN is higher than that for PT.

Similar to the observed data, the predicted detection threshold for the square wave is much higher than those for the other five stimuli. An examination of the power spectrum of the square wave shows that most of the energy is at the extreme low-frequency region where middle ear transmission is least efficient. Combining this with the same line of reasoning given for MPN, the skewed power spectrum has to have very high overall SPL in order to be detectable.

7.1.2 Loudness Functions

In this section, the predicted loudness functions for the six stimuli used in the loudness matching experiment are presented. However, in the following presentation, only the predicted loudness functions for PT, MPN, SP, and SQW are shown. The predicted loudness functions for narrow-band noises, NBN150 and NBN120, are not presented because of their close similarity to that for 1-kHz pure tone. The EIH-LM model used to predict the loudness functions are EIH-LM3 to EIH-LM12.

In Fig. 7.1, the loudness functions of the four stimuli predicted by EIH-LM3, EIH-LM4, EIH-LM6, and EIH-LM7 are compared to the observed loudness function for 1-kHz pure tone and a wide band noise (WBN). The observed loudness function for WBN is used to compare with the predicted loudness function because the observed loudness data for SP, MPN, and SQW are not available, and furthermore, it is to be expected that the observed loudness function for these stimuli would be similar to that for WBN, especially concerning the saturation problem. The loudness functions predicted by EIH-LM5 are not presented because they are essentially the same as those predicted by EIH-LM4 and EIH-LM6, except their saturation thresholds fall between those of EIH-LM4 and EIH-LM6. The predicted

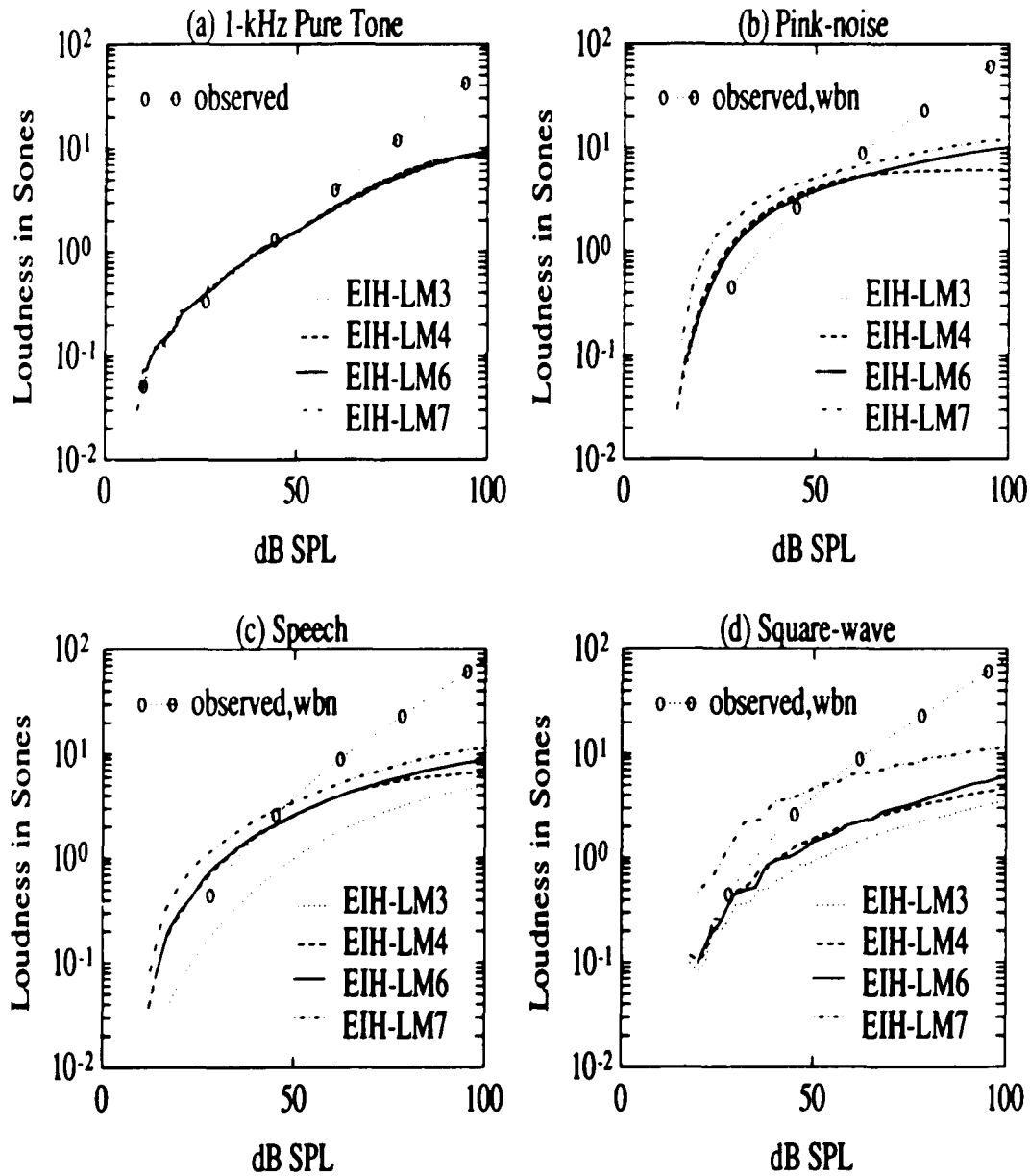


Figure 7.1: Loudness Functions Predicted by EIH-LM3, EIH-LM4, EIH-LM6, and EIH-LM7. Loudness functions predicted by EIH-LM3, EIH-LM4, EIH-LM6 and EIH-LM7 are shown. Also shown are observed loudness functions for 1 kHz tone and wide band noise, since the respective observed loudness functions are not available. The abscissa is the sound pressure level of the stimulus in dB SPL. The ordinate is the predicted or observed loudness in sones.

results for EIH-LM8 are also not shown because they are identical to those predicted by EIH-LM7.

The comparison between EIH-LM3 and EIH-LM4 shows the effect of a difference in the definition of predicted overall loudness, i.e. the mean instantaneous loudness over the signal duration vs. the 90th percentile of the instantaneous loudness distribution. As shown in panels (a) and (b) of Fig. 7.1, the two models provide similar predictions for pure tone and pink noise, but rather different predictions for speech and square wave, as shown in panels (c) and (d). The difference is relatively large for speech because of the differences between the mean and the 90th percentiles in the level distribution of a time-varying signal. The 90th percentile is affected less by low level components of the signal (including brief silences) than the mean of the distribution. The difference of the predictions by the two models for square wave is less than for speech, however, the same reason applies to the cause of the difference. In general, the differences between predictions of the two models vary widely depending on the amount of temporal fluctuation in the stimulus.

The loudness functions predicted by EIH-LM6 (the solid lines) for the four stimuli are essentially the same as those predicted by EIH-LM4 and EIH-LM5, except their saturation thresholds have increased to above 100 dB SPL. Thus the dynamic range of the predicted loudness is increased, an improvement by the EIH-LM6 model over previous model versions. However, the slopes between 40 and 100 dB SPL are still much too shallow to match those of the observed functions, indicating that any more modifications along the same line of EIH-LM4 and EIH-LM5 will not increase the slopes.

The effect of the removal of temporal integration rule 3 in EIH-LM7 on the predicted

loudness functions varies widely depending on the stimulus. The loudness function predicted for PT by EIH-LM7 remains the same as that predicted by previous versions of the EIH-LM model, while the predicted loudness function for SQW is shifted upward along the ordinate by half of an order of magnitude, quite large. The effect on MPN and SP is moderate. The resulting predicted loudness functions for MPN, SP, and SQW became rather similar in shape, saturation thresholds, and saturating loudness values. At low SPL, the three predicted functions grew much more rapidly than the observed loudness function for WBN (open circles connected by the dotted line), and much slower from mid- to high-SPL. It was also observed that after the removal of the temporal integration rule 3, the two ways of defining the overall predicted loudness, i.e. the mean and 90 percentile of the predicted instantaneous loudness, did not differ in any apparent way in the predicted loudness functions for complex stimuli except for speech.

The loudness functions predicted by EIH-LM8 and EIH-LM7 are identical indicating that increasing density of the level detectors within a fixed sensitivity threshold range with logarithmically equal spacing between the level detectors does not and will not change the slopes of the predicted functions.

The modifications in the EIH-LM model versions 9 to 12 mainly manipulate the spacing between the level detectors in more drastic ways in order to examine the effect of spacings on the predicted loudness functions. The predicted results by EIH-LM9, EIH-LM11, and EIH-LM12 are shown in Fig. 7.2 along with those predicted by EIH-LM7. The loudness functions predicted by EIH-LM9 and EIH-LM12 represent two extreme cases of the effect of the spacings between the level detectors in that the slopes of the predicted functions by EIH-LM9 are much steeper than those by EIH-LM7, while the slopes of the functions

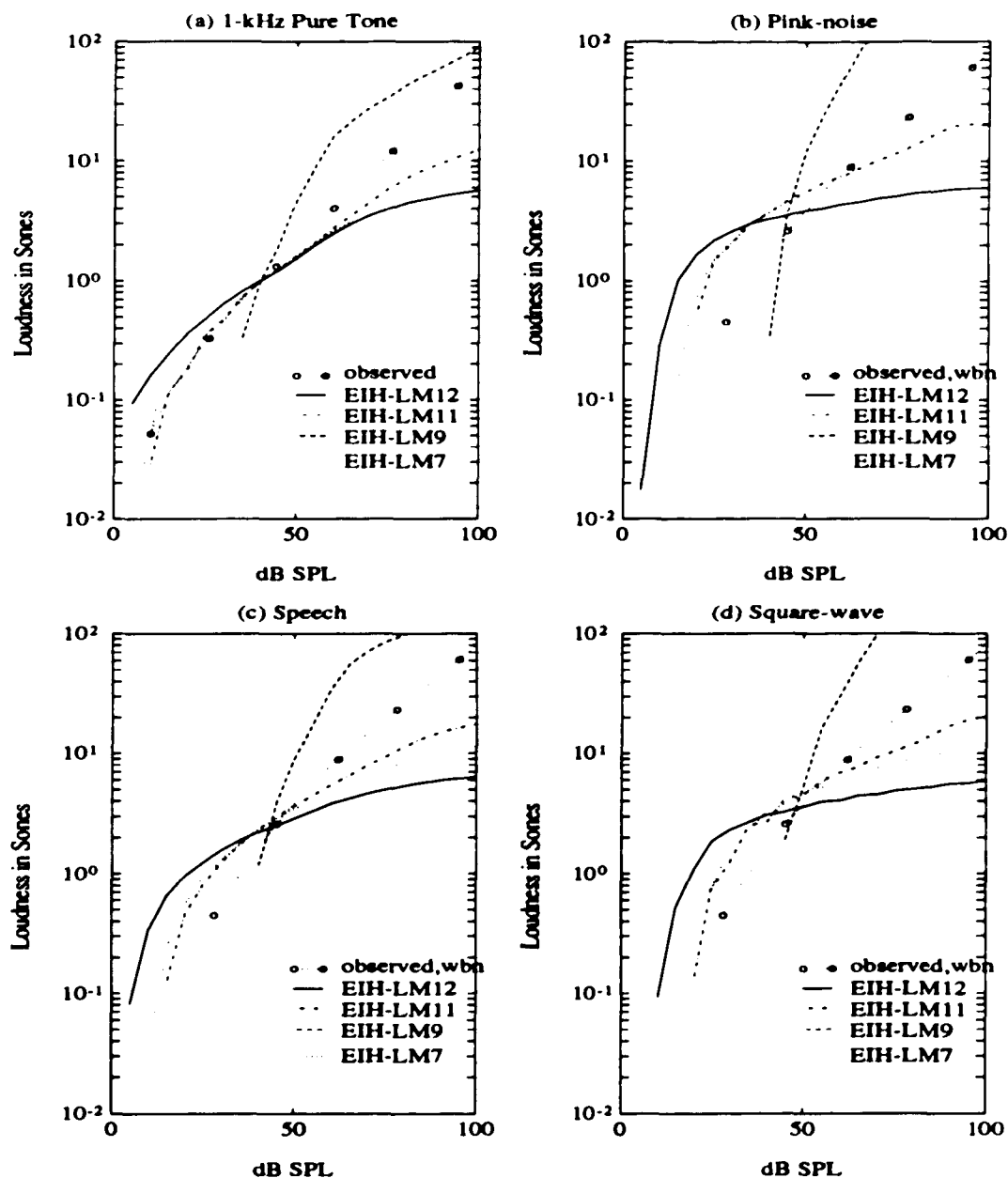


Figure 7.2: Loudness Functions Predicted by EIH-LM7, EIH-LM9, EIH-LM11, and EIH-LM12. Loudness functions predicted by EIH-LM7, EIH-LM9, EIH-LM11 and EIH-LM12 are shown. Also shown are observed loudness functions for 1 kHz tone and wide band noise. Since the respective observed loudness functions are not available. The abscissa is the sound pressure level of the stimulus in dB SPL. The ordinate is the predicted or observed loudness in sones.

predicted by EIH-LM12 are much shallower than those by EIH-LM7. In each panel, all the predicted loudness functions pivot (cross each other) at approximately the same area. Furthermore, the saturation thresholds decrease with decreasing slopes. At low SPL, all predicted loudness functions decrease very rapidly with decreasing SPL. The predicted loudness functions for EIH-LM10 and EIH-LM11 are not shown because the modification in EIH-LM10 is similar to that in EIH-LM9 and the modification in EIH-LM11 is a very mild case of EIH-LM9. As expected, the loudness function predicted by EIH-LM10 is very similar to that by EIH-LM9, while the loudness function predicted by EIH-LM11 is similar to EIH-LM7, except its slope above 50 dB SPL is steeper.

Discussion

1-kHz Pure Tone Saturation of the predicted loudness functions at the high sound pressure levels shown in panel (a) of Fig. 7.1, has been discussed in Chapter 5 in term of neural excitation patterns and the spread of excitation. In the EIH-LM model, the neural excitation pattern depends on the number of cochlear filters activated, on the number of level detectors activated, and the number of times each level detector activated over temporal integration time. In versions 3 to 6, the highest CF of cochlear filters is 12 kHz, which is an artificial limitation due to the lack of physical data and signal processing capacity. In cat, the upper frequency of the range of hearing is about 50 kHz and in human, it is no more than 20 kHz. Both of these numbers are much higher than the model's high frequency limit. If the model were expanded to incorporate more cochlear filters of higher CF, the predicted loudness function should saturate at a higher level, as was shown in Chapter 5 when the filter set was extended from 7.2 kHz to 12 kHz.

Other parameters, such as the number of threshold levels, the range of the total threshold levels, and the way they are distributed, can also affect the saturation level of the predicted loudness functions. The modifications in EIH-LM5 and EIH-LM6 are along this line. The saturation level of the loudness function predicted by EIH-LM6 is an improvement over the saturation level predicted by EIH-LM5, and that by EIH-LM5 is an improvement over that by EIH-LM4.

Pink Noise Comparing the predicted loudness functions of the MPN in (b) to those of the 1-kHz pure tone in (a) of Fig. 7.1, three points can be made: (1) the “detection threshold” of the pink noise is higher (discussed in Section 7.1.1); (2) the loudness growth rate of the pink noise is much faster at low to mid SPL range; and (3) the predicted loudness functions of MPN saturate at lower dB SPL is and at lower predicted loudness values.

Above the predicted threshold, the predicted loudness for the MPN increases rapidly, much faster than for the 1-kHz tone. The sudden increase of loudness is because almost all the cochlear filters are recruited at once, and the number of level detectors activated also increases with the increasing amplitude in each filter. However, in EIH-LM4, the level detectors from the filters of higher CF’s contribute sooner and faster than the low. These high CF filters also stop contributing sooner at lower SPL, because of the decreasing threshold range with increasing CF of filters. As is shown in Fig. 4.3, the filter of the highest CF stops contributing at 20 dB above the filter’s CF threshold. Therefore, the EIH-LM4 loudness function for pink noise saturate at a lower sound pressure level than that of the tone or of the narrow-bands noise.

As mentioned in Chapter 4, the effect of temporal integration time is much more complicated for a complex sound than for a pure tone. The maximum limit of integration time varies not only with the CF of the filter but also with the level detector and with every period interval. The compounded complexity, due to the complex sound and the non-regularity of the filters, makes it difficult to visualize the temporal process in the model. However, we will attempt to understand the effect of varying temporal integration time on loudness summation of pink noise in the EIH-LM model (including all versions) for two simpler cases. At low sound pressure, a filter passes only components with frequencies near its CF. As the amplitude of the sound pressure increases, the filters start to pass frequency components that are higher or lower than their CF's. The effect can be viewed in two situations: (a) low-CF filters taking in more higher frequency components and (b) high-CF filters taking in more lower frequency components. In case (a), the low frequency components have much larger amplitudes than the high frequency components. The time wave form of this composition is hence dominated by the low frequency components and the high frequency components should have small effect because the thresholds of the level detectors are spaced relatively far apart. In case (b), the low frequency components have much smaller amplitudes compared to the high frequency components. The combination gives an amplitude-modulated signal that would cause the neuron groups with low threshold levels to respond to the high frequency components and the neuron groups with high threshold levels to respond to the low frequency components and distortion products. Both cases have an effect on the saturation of the predicted loudness functions: In case (a), the effect is likely to be across all level detectors; and in case (b), the effect is more on the high-threshold level detectors. The effect is that fewer than 20 excitations may contribute

to loudness summation for those level detectors based on the temporal integration rule for complex sounds described in Section 4.1. Hence the predicted loudness functions of pink noise saturate at a lower loudness value.

In (b) of Fig. 7.1, the predicted loudness functions for the modified pink noise have much steeper slopes than the observed loudness function for a wide band noise near threshold. Due to the differences in the stimuli, it is difficult to give a quantitative judgement about how inadequate the EIH-LM model is. Qualitatively, since the modified pink noise was constructed to provide approximately equal energy in each critical band up to 7.2 kHz, it was intended to provide approximately uniform excitation along the basilar membrane. A wide-band noise constructed using equal spectrum level for all frequency components would have less energy per critical band in the low frequency region and more energy per critical band in the high frequency region. Hence, it seems possible that the modified pink noise would be louder than the wide band noise of equal overall level, and the observed loudness function for the modified pink noise would be slightly above that of the wide band noise. However, below 50 dB SPL, the difference between the predicted loudness of the modified pink noise and the observed loudness of wide band noise are rather large. It seems necessary to have some kind of mutual-suppression mechanism that would cause each channel to reduce its loudness contribution due to the neighboring channels' excitation, and vice versa. Two-tone suppression may be a mechanism.

Speech In order to predict the loudness function for speech, it is necessary to know the power spectrum of speech in addition to the loudness definition and the behavior of the model. The averaged power spectrum of the speech used in the loudness matching experiment has the shape of a broad-band flat noise centered geometrically at 1-kHz with

3-kHz bandwidth (250 Hz to 3500 Hz). However, unlike all the other stimuli used in the loudness matching experiment, the instantaneous power spectrum of the speech does not stabilize to one form but changes continuously throughout the whole sentence. Based on the definition for overall loudness in EIH-LM4, EIH-LM5, and EIH-LM6, it is the stressed vowels that dominate the overall loudness of the sentence.

Since the bandwidths of instantaneous power spectra of the vowels lie in between those of the pink noise and the narrow band noise, the predicted loudness function of speech should also lie in between those of the tone and the pink noise. That is the case for these predicted functions. The shape of the predicted loudness function for speech is similar to that of the pink noise, but with shallower slope at low SPL region and steeper slope at high SPL region. Basically, the same discussion for pink noise applies to predicted loudness functions for speech.

Square Wave The loudness functions of the square wave predicted by EIH-LM3 to EIH-LM6 show a high detection threshold and a very shallow growth of loudness, shallower than all the other stimuli above 50 dB SPL. The high threshold matches the detection threshold of the SQW measured in the loudness matching experiment. The extremely slow growth of predicted loudness in panel (d) of Fig. 7.1 is, as shown by EIH-LM7, due to temporal integration rule 3 that affects the square wave strongly. Since the spectrum of the square wave shows prominent components at very low frequencies, the dominance of these frequencies in each channel causes the number of excitations per level detector to be much less than 20 and hence contribute little to loudness summation in the model. The loudness function predicted by EIH-LM7 has a form similar to that for the modified pink noise concerning rapid growth of loudness near threshold. This seems to strengthen the

need of a mutual suppression mechanism to reduce the summed loudness across channels.

7.1.3 Loudness Level Functions

In this section, EIH-LM4, EIH-LM6, and EIH-LM7 are used to predict loudness level data for two of the test conditions in the loudness matching experiment, quiet (Q) and overall-loudness (O). The observed loudness levels used for comparison are the averaged loudness levels as shown in Figures 6.14 through 6.16, where sensation level is defined re detection threshold of the respective stimulus (SL). The predicted loudness level functions are expressed in sensation levels (SL) re predicted detection thresholds of respective stimuli.

Quiet Condition

Loudness level functions predicted by EIH-LM4 for speech (SP), narrow band noise (NBN150), and square wave (SQW), are shown in Fig. 7.3. Also shown are observed loudness levels when the test stimulus was adjusted (*) and when the reference stimulus was adjusted (o). The three panels in the left column show the results obtained using a 1-kHz pure tone as the reference. The panels in the right column show the results for narrow-band noise (NBN120) as the reference stimulus. The vertical and horizontal dotted line segments through the symbols show the standard deviations of the measurements.

The predicted loudness level functions in the panels on the left side are very similar to those on the right side, as are the observed loudness levels. This is to be expected since the predicted loudness functions for PT and NBN120 are very similar as indicated previously. A comparison of the predicted loudness level functions with the observed loudness level data show excellent agreement for narrow band noise (NBN150) as the test stimulus (see Panels (c) and (d).) Note that NBN150 is judged to be slightly less loud than a 1-kHz tone

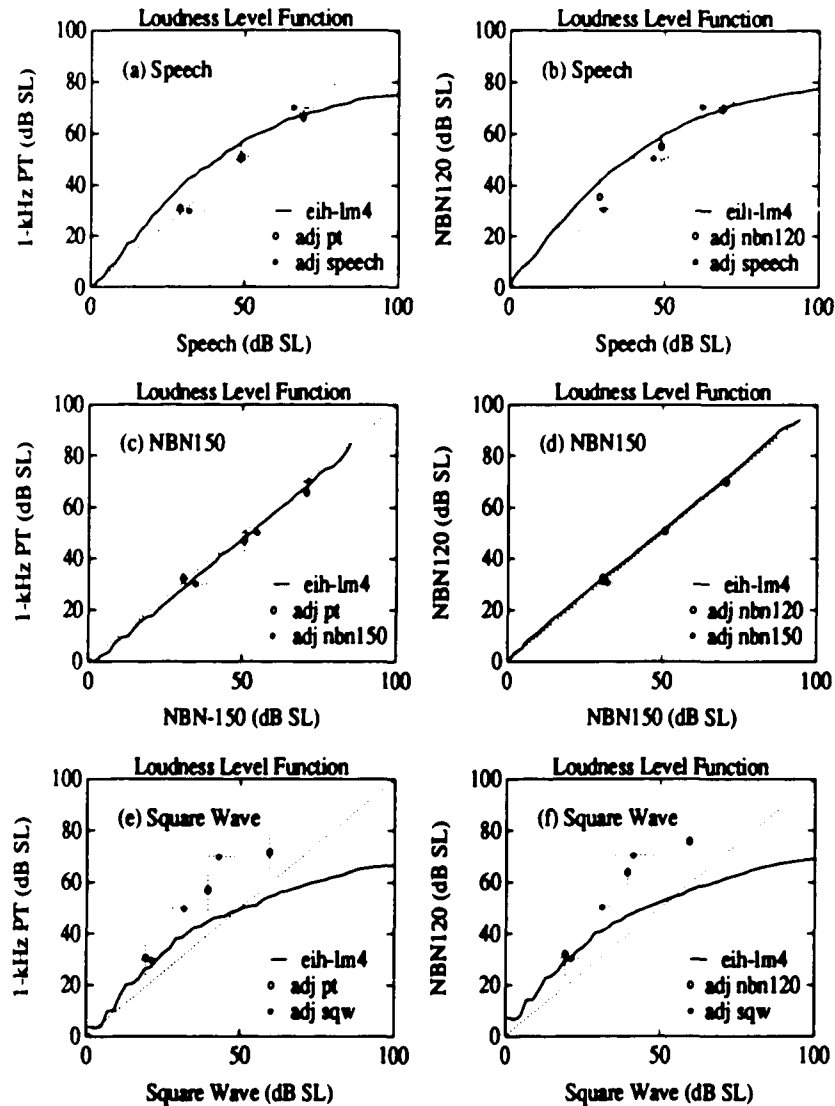


Figure 7.3: Predicted and Observed Loudness Level Functions in Quiet. Predictions are for the EIHLM4 model. Observed loudness levels are shown for the reference stimulus (o) and for the test stimulus (*) adjusted in level. The dotted line segment represent the standard deviations of the measurements. The left hand column of panels is for PT as the reference stimulus. The right hand column is for NBN120 as the reference stimulus. The top two panels ((a) and (b)) are for speech (SP) as the test stimulus. The middle two panels ((c) and (d)) are for narrow band noise (NBN150) as the test stimulus and the bottom two panels are for square wave (SQW) as the test stimulus. The diagonal dotted line in each panel represents the loudness level function for the reference stimulus.

of the same SL and that the predicted loudness level function show this difference quite clearly. According to critical-band experiments by Zwicker, Flottorp, and Stevens (1957), and Scharf (1959), the narrow-band noises, NBN150 and NBN120, should have identical loudness functions as the 1-kHz tone, because both narrow bands of noise have bandwidths less than the critical bandwidth centered around 1 kHz. In Chapter 5, we questioned the accuracy of that statement. Here we question it again.

The predicted loudness level function for speech does not differ from the observed loudness level data by a statistically significant amount; i.e., the predictions are adequate (in a statistical sense) for the range of loudness levels measured. Since the loudness level functions begin to saturate at levels above 50 dB SL, the quality of fit at high levels needs to be tested. Note that the diagonal line represents the loudness level function for the reference stimulus and that the predicted loudness level function for SP differs from this function. At low sensation levels (below about 35 dB SL), the predicted loudness level of speech increases faster than that of the reference (PT or NBN120). As the sensation level increases, the predicted loudness level function for speech increases at a slower rate than that of the reference stimulus. In contrast, the observed loudness levels appear to fall on the diagonal, indicating that the mean loudness levels for speech (averaged over subjects) is the same as that for 1-kHz pure tone (or NBN120). Examination of individual loudness level functions, however (see Fig. 6.2), shows that at least one subject has the same type of loudness level function for speech as that predicted by the model. Perhaps this type of loudness growth represents the tail of the distribution for normal-hearing subjects. Nevertheless, it is important to bear in mind the reason for the curvature of the predicted loudness level function for speech; at low sensation levels the predicted growth of loudness

for speech is faster than that at high sensation levels.

The predicted loudness level functions for the square wave (Panels (e) and (f)) fail to match the observed data. Below 50 dB SL, the observed loudness level functions, if fitted with a straight line, would have a slope greater than one, indicating that (1) the loudness of the square wave is greater than that of the reference tone at the same sensation level and (2) the rate of loudness growth for a square wave is greater than that of the reference tone stimulus (PT or NBN120). The predicted loudness level function of the square wave has a slope greater than one below 25 dB SL, but the function saturates rapidly above 50 dB SL, resulting in very poor predictions at high sensation levels.

One of the causes of the low saturation level of the predicted loudness level function for square wave is the third temporal integration rule which was suspected to have strong effect on complex sounds that have large low-frequency components. (See Chapter 4.1 for more details.) Another explanation is obtained by examining the power spectrum of the square wave, the cochlear filter set of the model, and the excitation pattern evoked by the square wave. The power spectrum of square wave shows that most of the energy is below 20 Hz. Above 20 Hz, the spectrum level rolls off at 20 dB per decade. However, the lowest characteristic frequency of the cochlear filter set in EIH-LM4 is 168 Hz. As a consequence, the spread of excitation toward low frequency with increasing sound pressure level is limited to 168 Hz, causing the excitation pattern of the model in the frequency region where the square wave has most energy to grow more slowly and to stop growing at lower level than in higher frequency regions. Thus the growth rate of the predicted loudness for square wave is much less than it would be if the cochlear filter set were extended to frequencies lower than 168 Hz, as in the human ear.

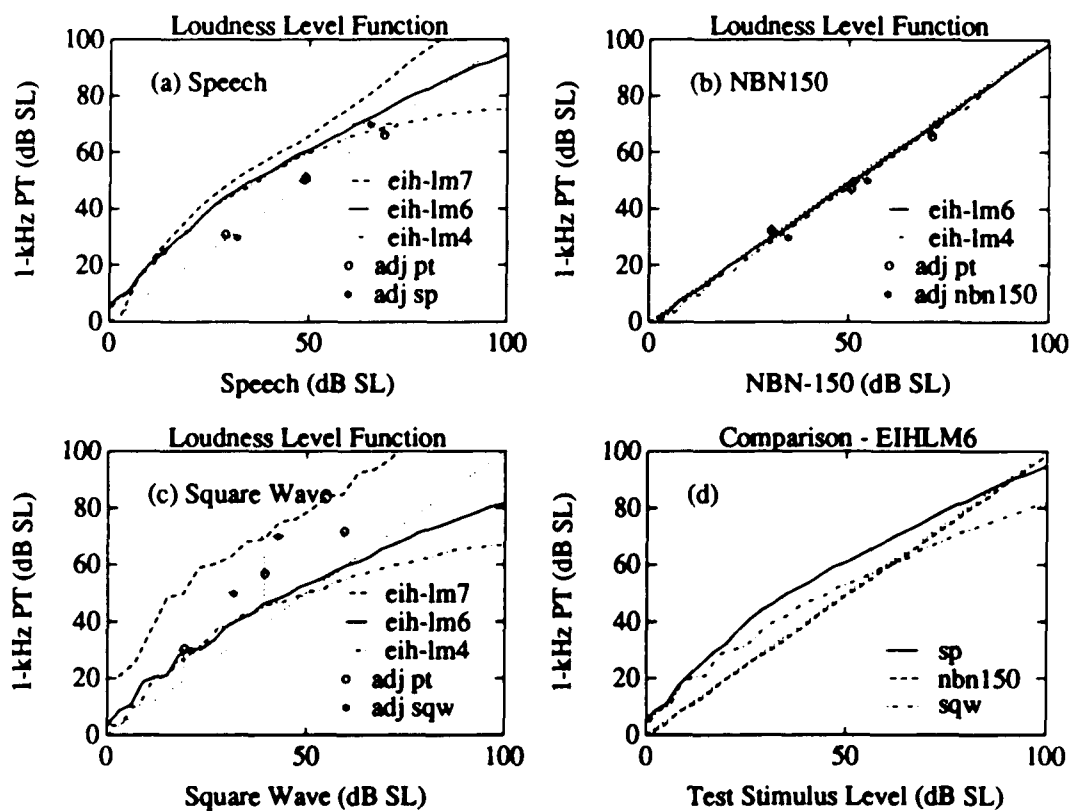


Figure 7.4: Comparison of Predicted Loudness Level Functions for EIH-LM4, EIH-LM6, and EIH-LM7 to the Observed Loudness Levels in the Loudness Matching Experiment. The format of each panel is similar to those in Fig. 7.3.

Comparing EIH-LM4 to EIH-LM6 Fig. 7.4 shows predicted loudness level functions for EIH-LM4 and EIH-LM6. The predicted loudness level functions for EIH-LM6 represent an improvement over those for EIH-LM4 because above 50 dB SL the slopes of the predicted functions of speech and square wave for EIH-LM6 are much steeper than those for EIH-LM4, hence the predicted functions are closer to the observed loudness levels. Consider the differences between the two models: in EIH-LM6, the number of the level detectors is increased from 5 to 8, the range of the detector thresholds is expanded from 60 to 80 dB, and the threshold range of the level detectors does not vary with the characteristic frequency of the filter. The net effect of the expanded range and increased number of level detectors makes EIH-LM6 more of a model with limited ability in simulating rate-level functions than EIH-LM4. This seems to suggest that neural rate-level functions are necessary in modeling loudness coding to cover a large intensity dynamic range. The loudness level functions for the three test stimuli predicted by EIH-LM6 are compared to each other as shown in panel (d).

Also shown in Fig. 7.4 are the predicted loudness level function by EIH-LM7, however for speech and square wave only (Panels (a) and (c)). The predicted loudness level function for speech is outside of the range of one standard deviation from the observed mean loudness levels. If the mean predicted instantaneous loudness were used as the predicted overall loudness, then the predicted loudness level function for speech would match the observed loudness level data nearly perfectly, i.e. the predicted function for speech would overlap that for the reference stimulus (the dotted diagonal line) with the same slope.

The removal of the temporal integration rule 3 has the most effect on square wave, as can be observed in Panel (c) of Fig. 7.4. Although the predicted loudness level function

passes near one of the observed loudness level at mid SL and some subject may have the loudness level data similar to the predicted function, the shape of the function in general has failed at both low and high SL's.

Overall-loudness Condition

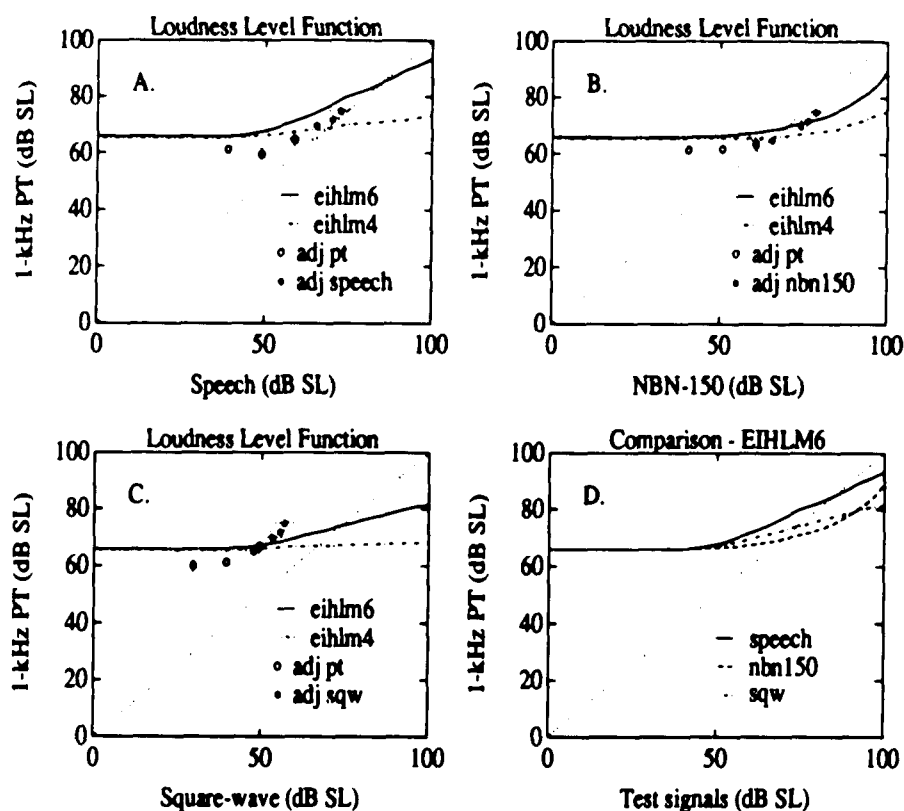


Figure 7.5: Comparisons Between the Predicted and Observed Loudness Level Functions for SP+PN, NBN150+PN, and SQW+PN. Loudness level functions predicted by EIH-LM6, by EIH-LM4, are compared to those measured using 1-kHz tone as the reference. The modified pink noise is set at a fixed level, 59 dB SPL. The format of this figure is the same as that of Fig. 7.4 except that the standard deviations are replaced by confidence intervals.

The loudness level functions predicted by EIH-LM4 for SP+PN, NBN150+PN, and SQW+PN using 1-kHz pure tone as the reference are almost identical to those using narrow band

noise (NBN120) as the reference stimulus. As a consequence, the predicted loudness levels shown in the section that follows are for 1 kHz tone only. Panels (a) through (c) in Fig. 7.5 show loudness level functions predicted by EIH-LM4 and EIH-LM6 for SP+MPN, NBN150+MPN, and SQW+MPN, respectively. Also shown are observed loudness level data using the same format as in Fig. 7.3. The improvement by EIH-LM6 at the high sensation level region is significant and is due to the same reasons explained for the previous condition.

In each panel, the observed loudness levels are shown in two segments, the flat segment (where the loudness of background noise (MPN) dominates) and the rising segment (where the loudness of the test signal dominates). For the flat segments, the loudness level functions predicted by both EIH-LM4 and EIH-LM6 lie above the mean of the observed loudness levels and slightly above the upper limit of confidence intervals of the means. The overestimation of the loudness for the flat segments is probably due to the faster growth of the predicted loudness of the modified pink noise relative to that of the 1-kHz pure tone at the low or mid-dB SPL. Perhaps, a suppressive mechanism involving mutual masking is necessary in the model to slow down the loudness growth of complex sounds, especially with wide band frequency components. The problem with the rising segment predicted by EIH-LM4 is due to the premature saturation of the pink noise at 60 dB SPL. Thus when the test stimuli with increasing amplitude is added to the modified pink noise, the predicted loudness is already in a saturated region and does not increase further.

For the rising segments, the loudness level functions predicted by EIH-LM6 reduce the saturation problem inherent in EIH-LM4 in two ways. Firstly, the slopes of these segments predicted by EIH-LM6 are steeper than those predicted by EIH-LM4, hence expanding

the dynamic range of the predicted functions. Secondly, the sensation levels at which the transitions from the flat segments to the rising segments occur correspond to those in the observed loudness level functions. Both improvements suggest that neural rate-level functions may be important for predicting loudness coding at high stimulus levels.

7.2 Zwicker's Method of Loudness Calculation

7.2.1 Implementation of the Method

The procedure used to calculate overall loudness, N , using Zwicker's method from a stimulus, $S(t)$, presented at eardrum is depicted by the block diagram in Figure 7.6. The first step of the procedure converts the input signal, $S(t)$, to a sequence of power spectra. The second step modifies the power spectra by Shaw's (1974) eardrum-to-free-field transformation function at 0-degree azimuth, as shown in Fig. 7.7. This step is necessary because the signals in the loudness matching experiment were presented at the subject's eardrum with the sound pressure levels calibrated at the eardrum while Zwicker's method for loudness calculation requires the third-octave-band measurements in free field as the input. The third step converts the modified power spectra into third-octave-band spectra. In the fourth step, the overall loudness, N , of the signal is calculated by Zwicker's method.

This last step is the heart of the calculations and consists of two steps. The first step converts the sequence of third-octave-band spectra into a sequence of "instantaneous loudness" values and the second step delivers one single value as the overall loudness of the signal based upon the statistics of all the instantaneous loudness values.

An instantaneous loudness value is the loudness value derived for one spectrum based

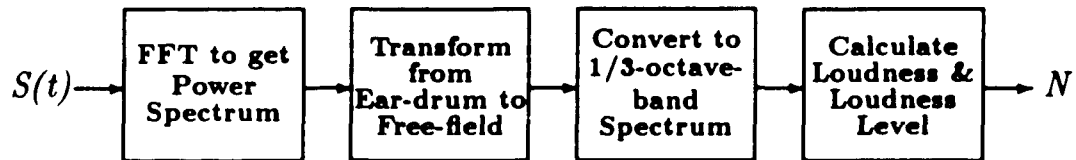


Figure 7.6: Block Diagram Depicting the Conversion Procedure Used to Convert Digitized Time Wave Forms of Sounds at Eardrum to Third-octave-band Spectra.

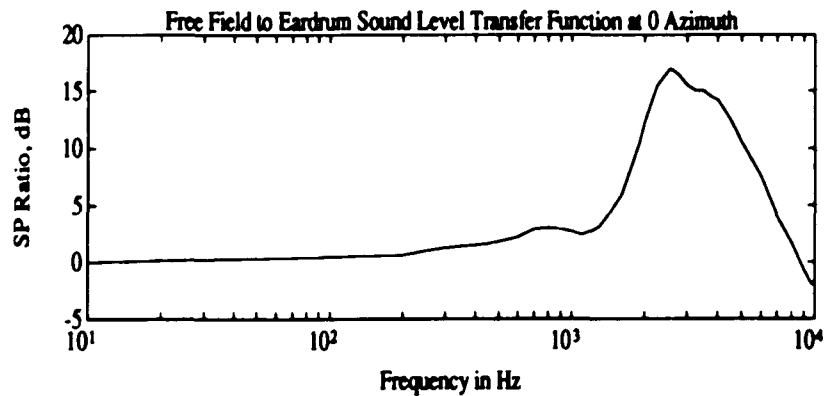


Figure 7.7: The Sound Pressure Level Transfer Function from Free-field to Eardrum at 0 Degree Azimuth. (Fig 9(a) in Shaw, 1974)

on a computer-program version of the ISO 532 B (Zwicker, Fastl, Dallmayr, 1984; Zwicker, 1985a). This method firstly converts the third-octave-band spectrum into a “specific loudness” diagram which represents the specific loudness per one-tenth bark as a function of one-tenth bark, from 1 to 240, then sums specific loudness across 240 one-tenth barks to give the instantaneous loudness in sones and the loudness level in phons. In order to distinguish between loudness and loudness level calculated using various methods, the loudness and loudness level calculated using Zwicker’s method are expressed in sones(GF) and phons(GF) as specified by ISO. The letter G in GF was used to distinguish the calculated sone from observed sone. The letter F in GF indicates the stimulus levels are measured in free-field. An example of a specific-loudness diagram for a speech spectrum is shown in the upper panel of Fig. 7.8 and the instantaneous loudness of the sentence in sone(Gf) as a function of time is shown in the lower panel of Fig. 7.8.

The overall loudness, N , over the entire sentence is defined to be the 95th percentile of all the instantaneous loudnesses. The 95th percentile point is derived using a histogram of the frequencies of instantaneous loudness. The value 95 is selected based on Zwicker (1977), Zwicker *et al.* (1985b), and Zwicker and Zwicker (1991). In the lower panel of Fig. 7.8, the overall loudness for the speech is 25.6 sone(Gf) as indicated by the dotted line across the peaks.

The window length used for FFT (Fast Fourier Transformations) is 51.2 msec for all loudness matching stimuli except speech, which is processed using 25.6 msec window. The FFT operation is done using a Hamming window at the rate of one every 25.6 msec, half of the window length, for all stimuli except speech, which is computed every 12.8 msec. In order to understand the effect of the FFT window length on the calculated loudness, the

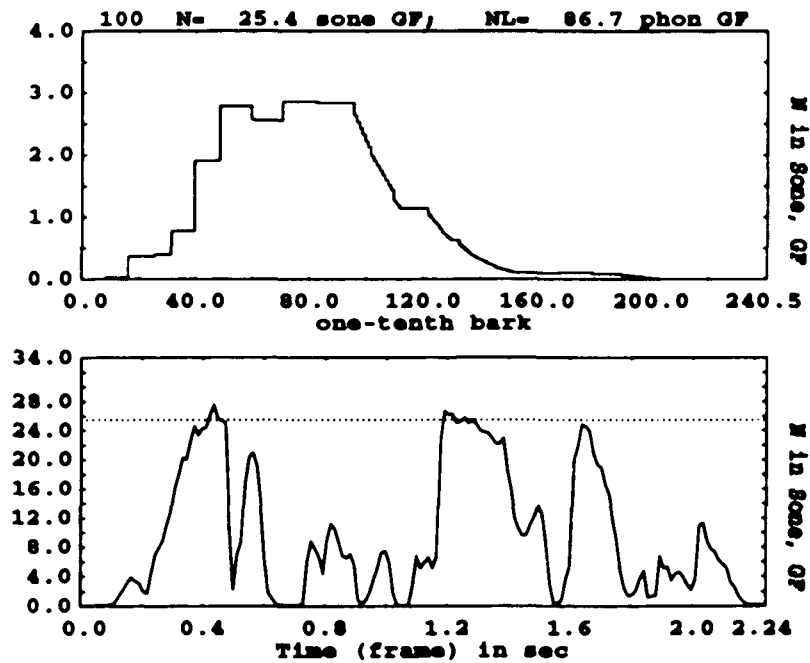


Figure 7.8: An Example of Specific-loudness Diagram of Speech and the Instantaneous Loudness of a Sentence as a Function of Time as Computed using Zwicker's Loudness Calculation Method. The upper panel is an example of specific-loudness diagram of a speech spectrum of 25 msec FFT window. The calculated loudness is 25.4 sone(Gf). The lower panel shows the instantaneous loudness of a sentence as a function of time in second. The overall loudness is 25.6 sone(Gf) using 95th percentile point of the instantaneous loudness as definition. The horizontal dotted line is at the loudness level where only 5% of the instantaneous loudness values exceed.

predicted loudness values of three representative stimuli are computed using six different FFT window lengths between 102.4 msec and 3.2 msec. The levels of the three stimuli are set at 80 dB SPL. The results are shown in Fig. 7.9, where the three stimuli are 1-kHz pure tone, speech, and square wave. All three stimuli show that except for the shortest window length at 3.2 msec, the predicted loudness decreases with increasing window length. The change is more apparent in the case of square wave, which stabilizes at 51.2 msec. For speech, the predicted loudness stabilizes at 25.6 msec, and for 1-kHz pure tone at 12.5 msec. Based on these results, 51.2 msec is selected to be the FFT window length to compute the power spectrum for all the stimuli except for speech (25.6 msec).

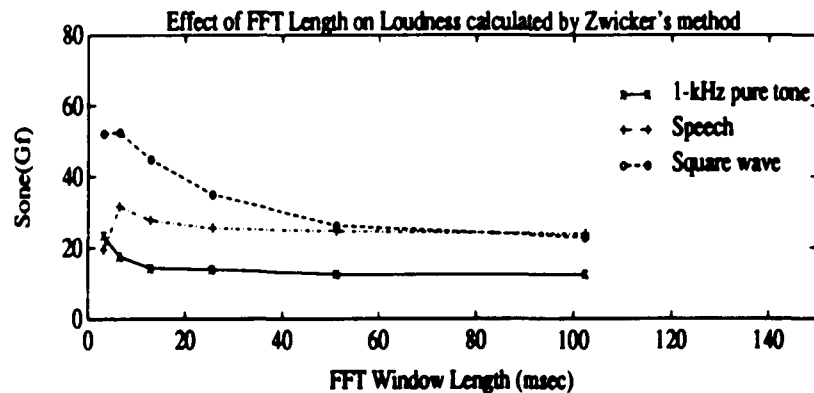


Figure 7.9: Study of the Effect of FFT Window Length on the Loudness Calculated by Zwicker's Method. The overall loudness of three stimuli are calculated for 6 different FFT window lengths, from 3.2 msec to 125 msec. The three stimuli are square-wave, speech, and 1-kHz pure tone at 80 dB SPL.

7.2.2 Loudness Function

Five loudness functions calculated using Zwicker's method are presented in Fig. 7.10. The five stimuli are speech (SP), narrow band noise (NBN150), 1-kHz pure tone (PT), square wave (SQW), and modified pink noise (MPN). The observed loudness function for 1-kHz pure tone is used to compare with the predicted functions for 1-kHz tone and for NBN150.

The observed loudness function for wide band noise is used to compare with loudness functions predicted for SP, SQW, and MPN.

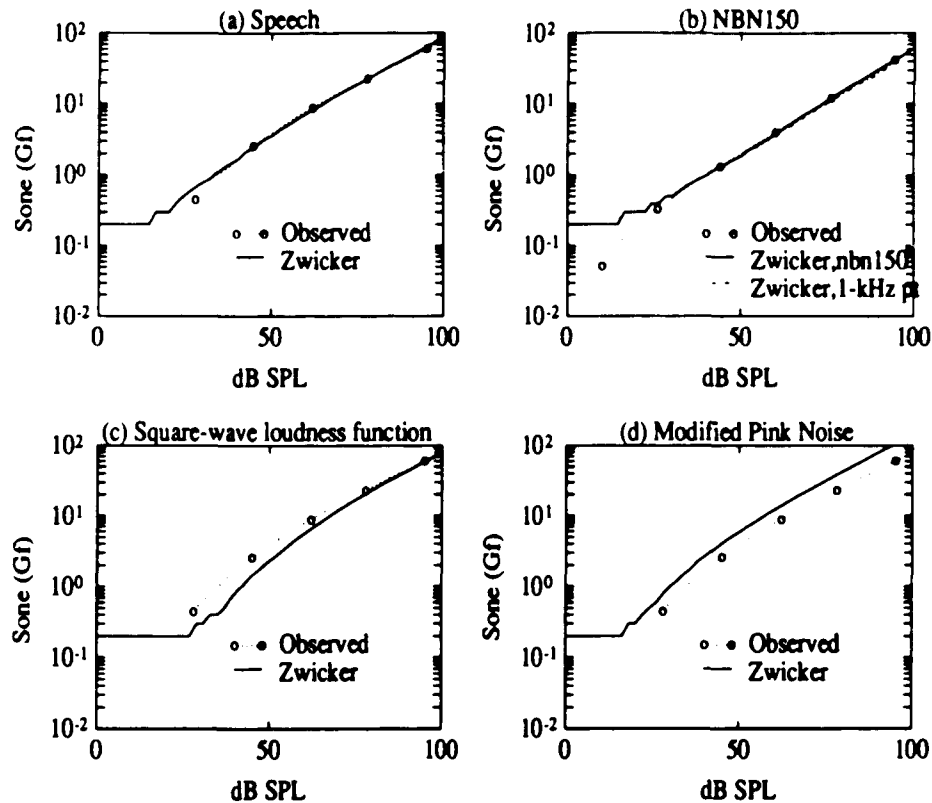


Figure 7.10: Loudness Functions Calculated Using Zwicker's Method for Five Stimuli Used in the loudness matching experiment. The five stimuli are speech in (a), narrow band noise and 1-kHz both in (b), square wave in (c), and pink noise in (d). The predicted loudness functions for NBN150 and 1-kHz pure tone are presented in the same panel for the purpose of comparison. Since the bandwidth of the narrow band noise (NBN150) centered at 1 kHz is 150 Hz, less than the corresponding critical bandwidth estimated by Zwicker, its predicted loudness is expected to be similar to that of 1-kHz tone. Also presented in (b) is the observed loudness function for 1-kHz pure tone. In (a), (c), and (d), the observed loudness function for the wide band noise is presented for a guide only, since the observed loudness function of speech, square wave, and modified pink noise are not available.

In panel (a), the comparison shows that above 35 dB SPL, the predicted loudness function for speech coincides with the observed loudness function for wide band noise. Panel (b) shows that above 30 dB SPL, the predicted loudness functions for 1-kHz pure tone matches that for NBN150, and both predicted functions match the observed loudness

function for 1-kHz tone. Panel (c) shows that above about 35 dB SPL, the predicted loudness function for square wave has a steeper slope than the observed loudness function of wide band noise. Panel (d) shows that the loudness function of pink noise computed by the model parallels that of wide band noise measured in humans. If the model prediction for the pink noise is correct, then the pink noise should sound louder than the wide band noise at the same sound pressure level. The implication is that at a fixed sound pressure level, the stimulus that excites the cochlear partition uniformly should sound the loudest, i.e. louder than sounds of other spectrum compositions. At the regions near thresholds, Zwicker's method gives a limit of 0.2 sone(Gf), not lower, hence all the predicted loudness levels have flat segments.

7.2.3 Loudness Level Functions

Quiet Condition

Two sets of loudness level functions for SP, NBN150, and SQW in quiet condition are calculated using Zwicker's method, as presented in Fig. 7.11. The first set is obtained using 1-kHz pure tone (PT) as reference (left column) and the second set is obtained using narrow band noise with 120 Hz bandwidth (NBN120) as reference (right column). Also presented are the loudness level data (o and *) for the three stimuli observed in loudness matching experiment.

Comparisons of the two sets of loudness level functions show that those obtained using NBN120 appear to match the observed loudness level data better. The model seems to underestimate the loudness of the 1-kHz tone so that it requires higher sensation level than humans in order to sound equally loud as speech. In panel (b), the loudness level functions

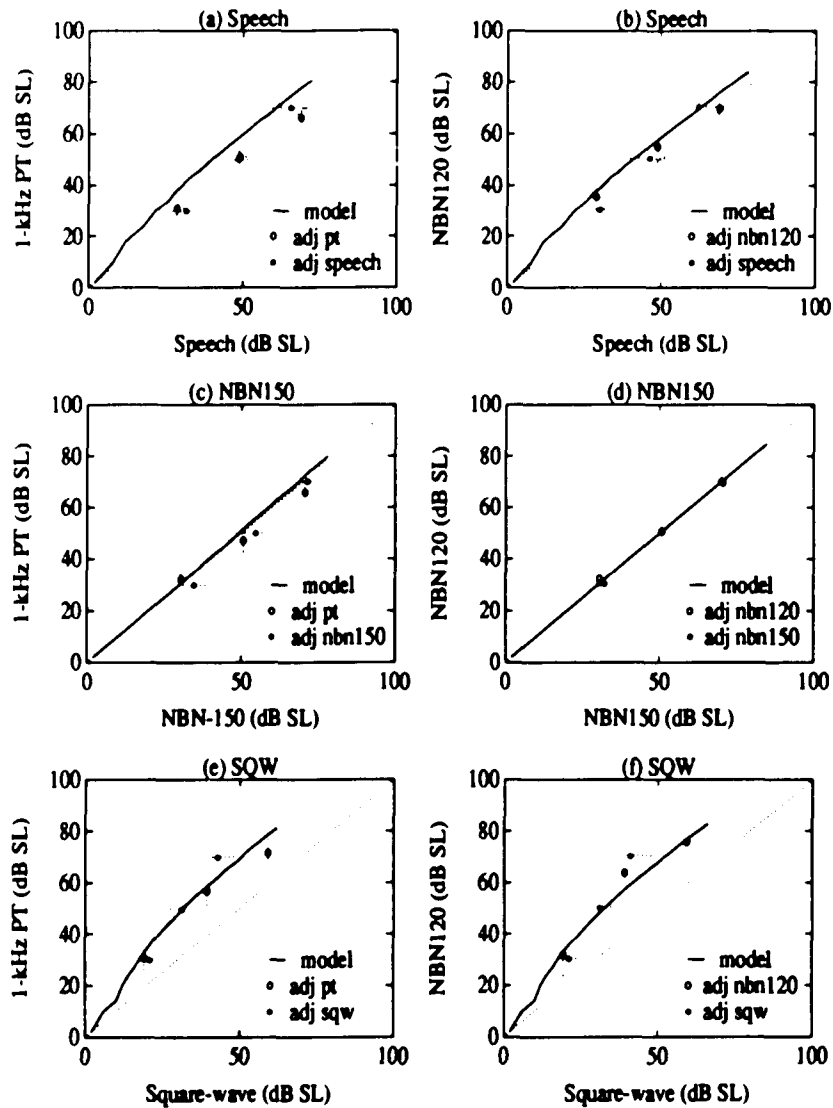


Figure 7.11: Comparisons of Loudness Level Functions Computed Using Zwicker's Methods with Those Observed. There are two sets of loudness level data for speech, narrow band noise (NBN150), and square wave (SQW). The first set is obtained using 1 kHz pure tone as the reference. The second set is obtained using narrow band noise (NBN120) as the reference. The symbols represent the observed loudness level data obtained in the loudness matching experiment.

calculated by Zwicker's method seems to have a steeper slope than the observed loudness level data. The predictions for the square wave using the two references have very similar results.

Partial-loudness (P) Condition

Method In this test condition, subjects are required to match the loudness of the test stimuli partially masked by a background noise to the loudness of the reference stimuli. To simulate this test condition, Zwicker's method of calculating loudness for partially masked stimuli (Zwicker, 1987) is used. This method uses the previous method to compute the specific loudness in one-tenth barks for both the masker and the maskee. In the loudness matching experiment, the masker is modified pink noise, and the maskees are the test stimuli. Then a set of three formulae is used to determine the partially masked specific loudness in each band based upon the specific loudness of the masker and the maskee in the corresponding band.

Following are the three formulae used to determine the partially masked specific loudness N'_p .

$$N'_p = 0, \quad \text{for } L_{TOBT} - L_{TOBM} < -4dB, \text{ i.e., } N'_T < 0.8 * N'_M$$

$$N'_p = N'_T, \quad \text{for } L_{TOBT} - L_{TOBM} > 12dB, \text{ i.e., } N'_T > 2 * N'_M$$

$$N'_p = (N'_T/0.6) - (8 * N'_M/6), \quad \text{for } 0.8 * N'_M \leq N'_T \leq 2 * N'_M$$

where L_{TOBT} stands for the third-octave-band levels of the test sound (T) and L_{TOBM} for those of the Masker. The test sound T referred to by Zwicker is the maskee mentioned earlier.

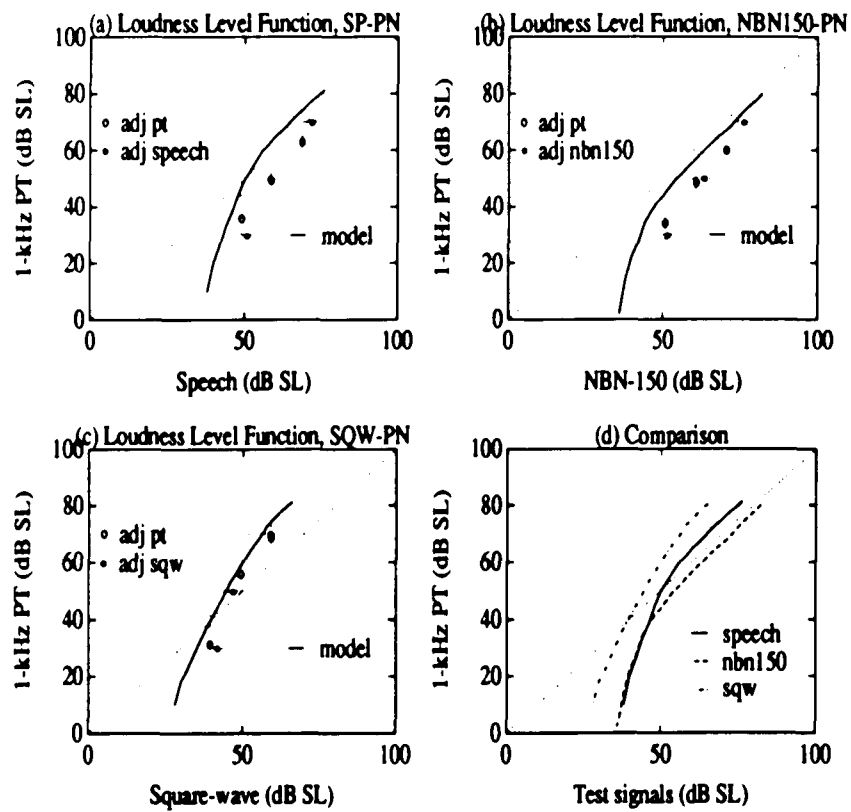


Figure 7.12: Comparisons of the Observed and Predicted Loudness Level Functions for Three Test Signals Partially Masked by the Background Pink Noise. The reference is 1-kHz pure tone. The symbols represent the observed loudness level data.

Prediction results The predicted loudness level functions for the partially masked test stimuli are presented in Fig. 7.12. The partially masked stimuli are represented as SP-PN, NBN150-PN, and SQW-PN. All three predicted loudness level functions in (a - c) have the right form, i.e. raised detection thresholds and rapid growth of the loudness above threshold, until the predicted loudness of the test stimulus in noise is the same as that of the test stimulus in quiet. However, the predicted loudness level functions for SP-PN and NBN150-PN show that in both cases the method overestimates the loudness of the partially masked SP and NBN150 (or underestimates the loudness of the 1-kHz pure tone). In the case of SQW-PN, the slope of the loudness level function at low sound pressure levels is shallower than that indicated by the observed loudness level data. The loudness level functions predicted for the two reference stimuli are very similar, hence only the set obtained using 1-kHz is presented.

7.2.4 Overall-loudness (O) Condition

In this condition, the loudness level functions for the sum of the test stimuli and modified pink noise are computed using the regular method, where the input are SP+PN, NBN150+PN, and SQW+PN. The resulting loudness level functions are presented in Fig. 7.13. Panels (a - c) shows that for all three test stimuli, Zwicker's method slightly overestimates both the loudness of modified pink noise in the flat segments and the overall loudness in the rising segments as the case in previous test condition. However, the forms of the functions are correct.

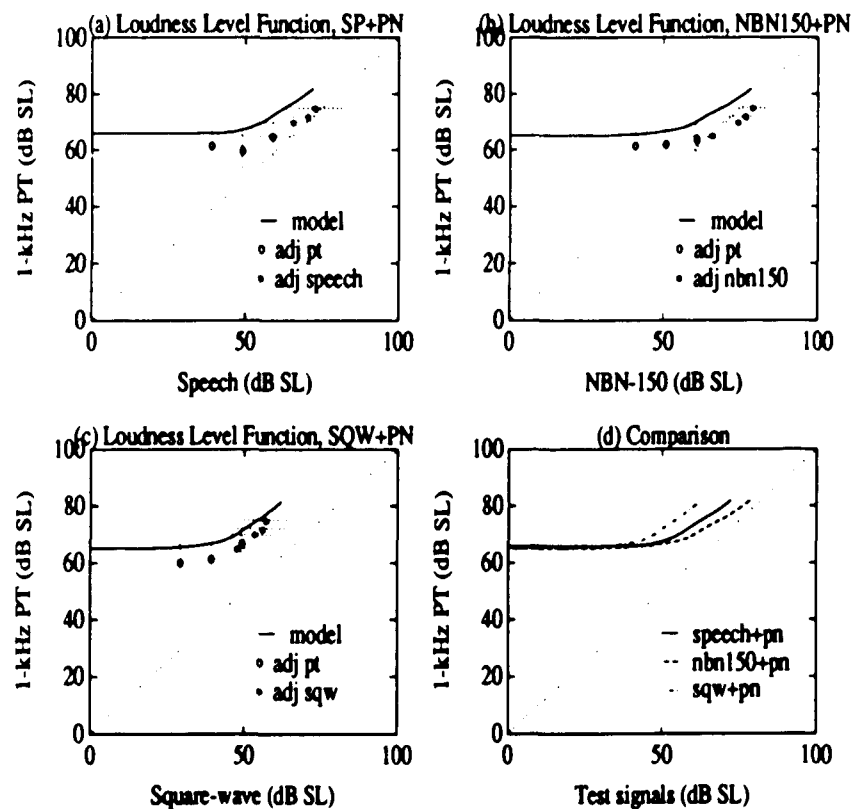


Figure 7.13: Comparisons of predicted and Observed Loudness Levels Functions for the Overall Loudness of Three Test Stimuli Plus the Background Pink Noise. This set of functions is obtained using 1-kHz pure tone and the format of this figure is the same as Fig. 7.11.

7.2.5 Discussion

In view of loudness predictions by Zwicker's methods in all three experimental conditions, all predicted loudness level functions have correct forms. However, one observation becomes apparent that is his methods either underestimate the loudness of 1-kHz pure tone or overestimate the loudness of most of other stimuli. The former seems more likely than the latter.

Chapter 8

Discussion and Conclusion

In this dissertation, a loudness model (EIH-LM), based on a physiologically based auditory model (EIH), was used to predict the loudness of pure tones and complex stimuli. In the process of loudness predictions, various model parameters were modified to (1) examine the effect of these parameters on the loudness predictions, and (2) evaluate the importance of the physiological mechanisms approximated by these parameters in the model. In addition, a loudness matching experiment was performed as part of this dissertation to obtain the loudness level data for three complex stimuli in three experimental conditions. The observed loudness data from two conditions were then used for loudness predictions by the EIH-LM models.

In order to provide a background for understanding the loudness model, the peripheral auditory physiology and some nonlinear phenomena simulated by the EIH model were firstly reviewed. Then, the loudness prediction methods developed by Fletcher, *et al.* at the Bell Laboratories since 1920's, by Stevens since 1950's, and by Zwicker, *et al.* since 1950's, were reviewed.

The loudness model used in this dissertation differs from the above-mentioned methods

of loudness prediction in that it is based on physiological functions believed to be basic for sound processing in the auditory system. In contrast, the loudness prediction methods of Fletcher, Stevens, and Zwicker are phenomenologically based; i.e. they are based on psychophysical observations and hence may be regarded as essentially interpolative of previously obtained loudness measurements. As such, they are useful as predictions but do not provide very much insight as to how loudness is coded in the auditory system.

Based on the view that the peripheral auditory system is the main source of most of the psychophysical observations, the loudness model consists mostly of the functions in the peripheral auditory system. The physiological functions simulated by the EIH model include the transfer function of the middle ear, frequency analysis of the basilar membrane, transduction function of the inner hair cell, sensitivity thresholds of auditory nerves, frequency analysis of temporal synchrony, and some kind of temporal integration.

The family of EIH-based loudness models developed in this dissertation is based on the EIH model. This model was originally used as a front end for an automatic speech recognition system. In this dissertation, the model was extended to include prediction of loudness. This was done by modeling the coding of loudness in terms of the number of neuron groups excited across a segment of the basilar membrane and the number of times these excitations occur over the temporal integration period.

In order to evaluate the model and to examine the factors important in loudness predictions, various parameters of the loudness model were modified for various predictions. The parameters modified were (1) the temporal integration rules; (2) the frequency range of the basilar membrane over which loudness was summed; (3) the definition of loudness over a period of time; (4) the distribution, the range, and the number of sensitivity thresholds

of the simulated auditory nerves;

Loudness predictions for pure tones and complex stimuli were presented in three forms: predicted loudness functions (in sones), predicted loudness level functions (in phons), and predicted iso-loudness functions. The first two functions were obtained for both pure tones and complex stimuli, while iso-loudness functions were obtained for pure tones only. In the first form, the predicted loudness of the test stimulus was expressed in sones as a function of SPLs of the test stimulus. In the second form, the predicted loudness of the test stimulus was also expressed as a function of the SPL of the test stimulus, except in the SPL of equally loud reference stimulus. In the second form, the sound pressure level of a reference whose predicted loudness was equally loud as that of the test stimulus was expressed as a function of the sound pressure level of the test stimulus. In the third form, the sound pressure levels of pure tones, whose predicted loudness were equally loud as that of the reference at a reference level, were expressed as a function of the frequencies of tones.

The stimuli used for loudness predictions included pure tones, a four-tone complex, speech, two narrow-band noises, a square wave, and modified pink noise. The loudness data used for predicting loudnesses of pure tones were obtained from the published data from Fletcher and Munson (1933), and Scharf (1978); loudness data for the four-tone complex were obtained from Scharf (1959). Loudness level functions for complex stimuli were obtained from a loudness matching experiment performed as part of this dissertation.

The discussions and conclusions in the following are divided into five parts: (1) a review of loudness prediction methods developed by the Fletcher's group and the problems facing these methods, and those prediction methods developed by Stevens and Zwicker *et al.*, (2) the results of the loudness matching experiment, (3) loudness predictions by the EIH-based

loudness model, (4) limitations of the EIH-based loudness model, and (5) recommendations for further research.

8.1 Review of Loudness Prediction Methods

The review of the early history of loudness prediction revealed that the pioneering work by Fletcher's group laid important foundation for psychoacoustics and loudness calculation methods. Many basic ideas were discovered in the process of developing loudness calculation methods. Unfortunately, much of their work is not (understood and not) recognized today. Furthermore, their line of research has been largely ignored. The following is a brief discussion of work in loudness calculation by the Fletcher's group and a summary of problems remaining in the current loudness calculation methods.

Three major bodies of work on developing methods of loudness calculation since 1920 were reviewed in Chapter 3. However, work accomplished by the Fletcher's group and the insights into the loudness summation problem and solutions were rather impressive. They used psychophysical measurements to gain insight regarding the underlying physiological mechanisms. They also used physiological knowledge to develop hypotheses to guide their psychophysical experiments.

Fletcher *et al.* were the first to use the concept of loudness summation to sum the loudness contributed by individual frequency components. They found that the loudness summation was a nonlinear operation, i.e. loudness from each frequency component did not contribute equally to the overall loudness, and the loudness of a stimulus is not equal to the sum of the loudness of all individual component. They also found, as early as 1924, that using a band of frequencies rather than single frequency component as a unit for

loudness summation gave better results. Eventually, this realization led them to discover the critical band concept using masking experiments. Furthermore, they found that overall loudness was equal to the cube root of the loudness of individual frequency components at high sound pressure levels. The exponent increased with decreasing sound pressure of the stimulus, approaching unity near threshold. These findings appear to be the first realization of a power law relating the loudness of a sound to its sound pressure.

In search of a method to predict loudness of all types of stimuli, Fletcher *et al.* explored the loudness of pure tones of various frequencies (64 Hz to 16 kHz) at various sound pressure levels (threshold to 110 dB SPL) relative to the loudness of a 1-kHz tone. These measurements resulted in what is commonly known today as the Fletcher-Munson pure tone iso-loudness contours. Via this set of iso-loudness contours, they were able to relate the loudness of any single tone to that of a 1-kHz reference. It is interesting to note that whereas the contribution of Fletcher and his co-workers to the development of iso-loudness contours is well known, their equally important contribution showing a power law relationship between loudness and stimulus intensity is not widely recognized.

Based on their observations of loudness growth as a function of sound pressure level, they were the first to develop a loudness scale based on a hypothesis that loudness was coded by the composite firing rate of all auditory nerve fibers reaching the brain. The loudness scale expressed the loudness of a sound in loudness units proportional to the hypothesized total firing rate. If the loudness doubled, the number of loudness units doubled. A loudness function for a 1-kHz tone was derived expressing the loudness of the tone in loudness units as a function of sound pressure level. This loudness function showed a doubling of loudness for every 9 dB increase in sound pressure level above approximately

40 dB SPL, and a rapid reduction of loudness with decreasing sound pressure level below 40 dB until threshold (see Fig. 3.3.) This loudness scale appears to be the forerunner of currently accepted loudness functions. Unlike modern loudness functions, it was derived from psychophysical measurements based on a physiological model of the auditory system.

Using the iso-loudness contours and the 1-kHz loudness function Fletcher *et al.* developed a method for subdividing the power spectrum of a stimulus to a spectrum of loudnesses of individual components. However, due to mutual masking between neighboring components, only a fraction of the loudness from each component contributed toward the overall loudness. This spectrum of fractional loudness as a function of frequency region was viewed as a frequency-dependent and level-dependent neural excitation pattern of the stimulus. The sum of the fractional loudness across all components was then the calculated loudness of the stimulus. The determination of the fraction of loudness contributed by each component was found to be a difficult problem, since the reduction of loudness in individual components caused by masking of neighboring components depended on the distance between the regions of the basilar membrane excited by the two interfering components and their relative spectral levels. Various methods were explored to determine the fractions, including estimation of the excitation patterns using masking patterns of the stimuli and a function relating masking threshold of a component to the fractional loudness of the component under the masking effect of the neighboring components (Fletcher and Munson, 1938).

During the course of the above investigations, Fletcher and his co-workers generalized the method of loudness prediction for complex stimuli, and provided much insight to the hearing mechanism. As a result of their work, current models of loudness summation, even

those developed primarily from psychophysical data, depend on the concepts of critical bands, and the power law combined with observed masking data for narrowband stimuli; i.e., although not usually credited with laying framework of modern loudness models, almost all of these models depend on concepts introduced by Fletcher and his research team. Each of these concepts has problems.

Despite the importance of the concept of the critical band, there are problems with using critical bands which need to be recognized: (1) The frequency domain is divided into consecutive bands, while the frequency analysis property of the basilar membrane is an overlapping continuum. (2) The physiological origin of the critical bands are still not clear at present. (3) The conclusion by Zwicker *et al.* (1957) and by Scharf (1959) concerning the critical bandwidth measured using loudness seems questionable. The critical band data measured by Scharf, as discussed in Chapter 5, show that loudness of narrow band complex sounds does not necessarily remain the same with bandwidths less than corresponding critical bandwidth. Further, Scharf's data do not show clear-cut critical bands and, as discussed in Chapter 5, there are differences between what the data show and the reported bandwidths. An important additional observation is that the critical bandwidths do not seem to be independent of sound pressure level.

The problems with using masking patterns to estimate the excitation patterns are: (1) Masking patterns of pure tones, as demonstrated by Delgutte (1989, 1990), are more complicated than what was assumed, due to two tone-suppression and off-frequency listening (as discussed in the Discussion section of Chapter 3.) (2) Using masking patterns to predict excitation patterns of complex signals is even more difficult because of the interactions between components and nonlinear additivity of two different maskers.

Using a physical auditory model based on physiological data has its own complexity and difficulty. Nevertheless, it solves the critical band problems by using overlapping cochlear filters. The problem of masking is much more difficult. Only when the mechanisms responsible for masking are fully understood, will the main stumbling block for loudness prediction be removed.

8.2 Loudness Matching Experiment

The main objective of this experiment was to obtain loudness level functions for speech, narrow-band noise, and square wave, in three experimental conditions. In the first experimental condition (Q), the loudness level functions of the test stimuli in quiet were obtained. In the second experimental condition (P), the loudness level functions of the test stimuli partially masked by noise were obtained. In the third experimental condition (O), the loudness level functions of the test stimuli plus the background noise were obtained. In both conditions P and O, the background noise was a modified pink noise set at 50 dB SL1k, as described in Section 6.2.3.

A preliminary analysis showed that the variability of the data satisfied the requirement of homogeneity of variance required by traditional analysis of variance. This assumption was found to be valid within subjects but not between subjects and that between-subject differences were large. As a consequence, it is important that a repeated measures analysis be performed since this is the major source of variability.

The loudness level functions for the three experimental conditions were significantly different and each reflected the effect of tasks and of the background noise. Those obtained in P showed a recruitment effect, and those obtained in O showed a lower boundary because

of the background noise. The loudness level functions for the three test stimuli were significantly different for all three experimental conditions. However, the loudness level functions for speech (SP) and narrow-band noise (NBN150) were similar, while that for square wave (SQW) was very different.

A narrow-band noise (NBN120) was used as the reference stimulus as well as a 1-kHz pure tone (PT), because (1) according to critical band measurements by Zwicker *et al.* and by Scharf, NBN120 and PT should sound equally loud for equal SPL; and (2) since the test stimuli were very different from a 1-kHz pure tone, NBN120 might be easier to use than PT as a reference. The experimental results showed that there was no systematic or statistically significant test-retest variation due to the two types of reference. Subjects' comments also gave no clear-cut indication that one reference was favored over the other.

The difference between mean responses using the two references was found to be statistically insignificant. However, there was a systematic difference between the loudness level functions measured using the two references, as can be observed in Fig. 8.1. Across all conditions, and all test stimulus types, the loudness level functions measured using NBN120 lie above those measured using PT. Although the final loudness level functions were obtained by averaging over the loudness data measured using the two references, the variations due to the two references were important from the interpretation point of view. Therefore, the loudness level functions from the two references are treated separately in the following discussions. In Fig. 8.1, plots (a - c) are for condition Q, (d - f) for condition P, and (g - i) for condition O.

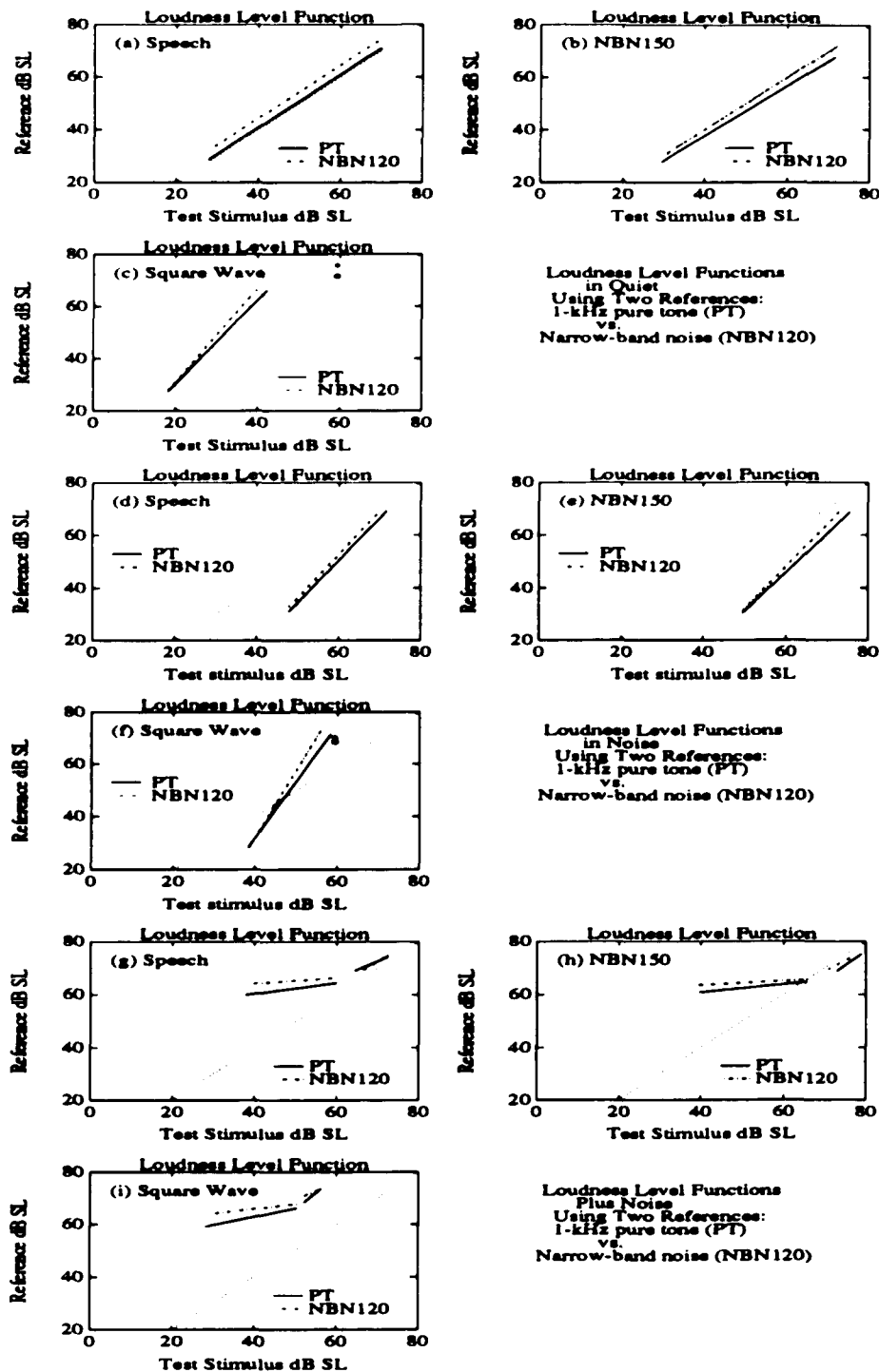


Figure 8.1: Loudness Level Functions Measured Using PT and NBN120 as the Reference Stimulus for Each of the Test Stimulus in Each of the Experimental Conditions. Panels (a-c) represent those measured for Q condition, panels (d-f) for P condition, and panels (g-i) for O condition.

8.2.1 Loudness Measured in Quiet (Q)

Speech and Narrow-band Noise (SP and NBN150)

Two pieces of evidence establish that the difference between the loudness level functions obtained using the two references was real. Firstly, referring to Fig. 8.1 panel (b), the function for NBN150 with NBN120 as the reference coincides with the dotted diagonal line, showing that NBN150 and NBN120 had the same loudness growth rate and sounded equally loud for the same SL. This was expected because the two stimuli were the same except the bandwidth for NBN150 was 30 Hz wider and both bandwidths were less than the critical bandwidth of the corresponding center frequency. Secondly, the loudness level function for NBN150 using PT showed that the loudness growth rate for NBN150 was slightly slower than that for PT, so that NBN150 sounded softer than PT at the same SL. This finding conflicts with the conclusion drawn by Zwicker, *et al.*, Scharf, and Fletcher that stimuli of equal energy should sound equally loud if their bandwidths are less than one critical band. However, as was shown in Section 5.2, our finding is not in disagreement with the earlier data. The finding persists throughout the measurements in this experiment. Loudness levels measured using NBN120 were consistently higher than those measured using PT, because at equal SL NBN120 was softer than PT.

The estimated loudness level function for SP using PT had a slope of 1.01 and an offset of only 0.6 dB, indicating that SP and the reference stimuli had identical loudness growth rates over a wide range of stimulus level (30 to 70 dB SL). This indicates that 1-kHz pure tone is a reasonable reference stimulus for representing the loudness of speech. It is common practice to use a calibration tone to specify speech levels in psychophysical experiments. This practice is not unreasonable given that it is easier to generate and measure. However,

one must caution that only one sentence was used in this experiment. It is commonly known that the detection threshold and the loudness of speech do not depend on the entire sentence but on a few stressed syllables. How a sentence composed of mostly consonants and another sentence composed of mostly vowels would differ in loudness growth rate is not clear. Further study to examine the effect on loudness perception of different proportions of vowels and consonants would be required in order to draw a general conclusion.

The loudness growth curve showed that SP sounded as loud as PT at the same sensation level. The measure of speech level used in this experiment was the 90-percentile-point SPL (90%-point SPL), the SPL that exceeded 90% of the instantaneous SPL. The speech level determined this way reflected the peak energies of the sentence. This method was found to be objective and stable. It was not sensitive to the variation of the FFT-window (Fast-Fourier Transformation) length in processing the speech signal. This was important for making the same experiment repeatable.

Square Wave (SQW)

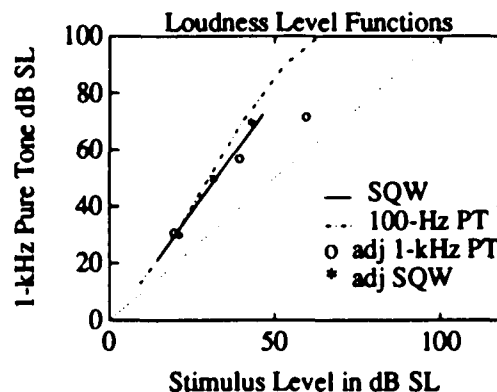


Figure 8.2: Loudness Level Function for SQW (the solid line fitted to the open circles and the asterisks) and for 100-Hz Pure Tone (dash-dotted line). The loudness level data for 100-Hz pure tone were obtained from Fletcher and Munson's paper in 1933.

The abrupt decrease of the loudness growth rate above 40 dB SL can be explained by examining the spectrum composition of SQW. The SQW spectrum was dominated by low frequency components. Above 60 Hz, the spectrum level rolled off at about 20 dB per decade. The results indicate that the loudness of SQW was also dominated by low-frequency components. To demonstrate this point, the loudness levels of 100-Hz pure tone were extracted from the equal-loudness contours of the pure tones by Fletcher and Munson (1933) and plotted in dB SL as the dash-dotted line in Fig. 8.2. The loudness level function of SQW is represented by the solid line and open circles and asterisks. The two lines coincide at low intensities. Between 10 to 40 dB SL, the function for 100-Hz pure tone has a slope of 2 and for SQW 1.6. The loudness growth rate for SQW is slower than that for 100-Hz tone. However the dominance of the low frequency components in SQW's loudness growth is clearly suggested.

The exact loudness growth pattern of SQW at high level could not be determined from this experiment. Another experiment with finer sampling of sound levels at high level would be necessary to provide a better picture. The SQW sounded like a jackhammer, which is considered industrial noise. In attempting to control environmental noise, various measuring methods have been used to estimate the amount of annoyance these noises generate. A-weighted sound pressure levels, loudness levels and, loudness in sones have all been used. The SQW data here showed that at low SL, one could easily underestimate its loudness. Further study of the loudness growth of this type of stimulus with finer intensity resolution is indicated.

8.2.2 Partial Loudness of Masked Stimuli (P Condition)

As shown in panels (d-f) of Fig. 8.1, the background noise caused the detection thresholds and the slopes of the loudness level functions of the partially masked test stimuli to increase. However, the loudness levels returned to normal at 70 dB SL for SP and NBN150 and at 60 dB SL for SQW. The effects of noise on both threshold and the slope were of similar magnitude for all three test stimuli. At 30 dB loudness level, the test stimuli levels had to be increased by 17 to 18 dB from the Q to the P condition. The slopes were increased by a factor of 1.56 for SP and NBN150 and 1.33 for SQW. This effect is similar to *recruitment* in hearing impaired listeners.

The above observation has an interesting audiological implication. If loudness recruitment in noise-induced hearing impairment is similar to that produced by masking in normal ears, and if the hearing loss is relatively constant or mildly sloping with frequency, then presenting compressed signals to hearing-impaired persons (1) should not distort the loudness ratio of various components in the signals and (2) should restore the normal loudness growth pattern. Furthermore, one compression ratio over the main hearing range would be sufficient for most stimuli, except for signals like SQW. At high SL, the loudness growth rate of SQW changed with the intensity. If the hearing loss were severe or if the background noise were higher in level, it is not clear how the loudness growth rate would change. Neither is it clear whether a variable compression ratio with increasing intensity would be necessary. However, the finding is encouraging. This finding in conjunction with the relationship between just-noticable difference in the impaired ear and the normal ear would allow for the development for models for predicting loudness growth in hearing impaired listeners.

8.2.3 Overall Loudness of Masked Stimuli (O Condition)

When the test stimulus was adjusted so as to match the total loudness of the test stimulus plus background noise to the reference, the first three subjects (S4, S5, and S7) had considerable difficulty for a 65 dB SL1k reference stimulus. Therefore the reference stimulus was increased to 72 dB SL1k for the rest of the subjects. Two explanations for the difficulty these subjects had are offered. From the lower segments of Fig. 8.1 (g - i), one might think that since the total loudness levels for the test stimulus at 40, 50, or 60 dB SL1k varied between 60 and 65 dB SL1k, then there should be no problem if the reference stimulus were at 65 dB SL1k. However, there was a problem, because of the regression effect discussed in Section 6.5.4. Because of the subjects' tendency to gravitate toward more comfortable levels, the actual loudness levels for the lower segment might be higher than 65 dB SL1k. Hence it would be impossible to find a test stimulus level to give a total loudness level of 65 dB SL1k.

The other explanation is based on subjects' comments and on numerous listening to these test conditions. There was a large range of test stimulus levels over which much trading occurred between the amount of the stimulus being masked and the loudness contributing to the total loudness. It seems likely for one to find a range of test stimulus levels to match the total loudness of the test stimulus plus MPN to the loudness of the reference stimulus at a loudness level. It also appears that the total loudness curve might be a non-monotonic function of the test stimulus level, i.e. as the test stimulus level was increased, the total loudness might decrease at some level, because the level was high enough for the test stimulus to mask out most of the MPN energy but not high enough to increase the total loudness. This phenomenon was also observed by Hellman (1991). However, when a

subject assumed that the loudness summation of a test stimulus plus MPN must be louder than MPN alone, then it became impossible to perform the loudness matching.

A Hypothesis The existence of two segments in the loudness level functions and the abrupt transition from a plateau to normal loudness growth, as in quiet, requires an explanation of how loudness of two stimuli are summed as a function of test stimulus level. Most of the subjects complained that among the three experimental conditions and the two adjustment methods, total loudness matching by adjusting the test stimulus was the most difficult task. To perform total loudness matching, subjects had to try to fuse two different stimuli together to listen for the total loudness, and then to match the total loudness to the loudness of the reference stimulus. The difficulty of fusing two sounds together seemed to vary with the degree of difference of the two sounds and their relative levels. Among the three test stimuli, SP, NBN150, and SQW, the most difficult to fuse with MPN seemed to be SP. One possible reason was that we have been trained to extract speech information out of background noise. At low level, SQW blended more easily with the MPN. NBN150 had strong tonal component which stood out in the MPN, and it was easier to identify.

Many subjects indicated that when it became impossible to fuse the test stimulus with the MPN, they chose the loudness of the dominant sound to be the total loudness. Assume this behavior of finding the total loudness to be true. That is, when the loudness level of the test stimulus was *lower* than that of the MPN, the subjects equated the loudness of the MPN to the total loudness, and when the loudness level of the test stimulus was *greater* than that of the MPN, the subjects equated the loudness of the test stimulus to be the total loudness. Then the upper segment of each loudness level function should approach the corresponding loudness level function in quiet. A comparison between the upper segments

of the loudness curves for condition T and the corresponding loudness curves for condition Q, as depicted in Fig. 6.17, indicates the hypothesis is approximately true. The deviation from the hypothesis might be due to the degree of the difference between the two stimuli being fused. One has to also bear in mind that the total loudness curves have not been corrected for the regression effect and the loudness level functions for SQW at high SL were not completely clear.

8.3 Loudness Predictions by the EIH-LM Model

The loudness predictions for the EIH-LM model were limited to the frequency range from 150 Hz to 4 kHz. The limit of the frequency range was due to the temporal synchrony simulated by the EIH model. However, the principle of the loudness defined for the EIH-LM model does not have to be restricted to the upper frequency limit of 4 kHz. The original EIH-LM model was used to predict the loudness of pure tones only. Then various parameters of the model were modified to examine the effects of these parameters on the predicted loudness for pure tones and complex stimuli and thus the model was evaluated. In this section, a summary of the loudness predictions by the original EIH-LM model is presented first, then the discussions of the effects of various parameters follow.

8.3.1 Summary of the Predictions by EIH-LM1

Loudness Function

(1) The model was able to predict pure tone loudness functions below 50 dB SPL for frequencies from 500 Hz to 3 kHz. The interesting point is that the shape of the observed loudness function for a 1-kHz pure tone can be divided into two sections, above and below 40 dB SPL. Above 40 dB SPL, the loudness function is a straight line in a log-log plot

and can be easily predicted by a power function. Below 40 dB SPL, the loudness decreases much rapidly with decreasing SPL and is more difficult to predict for most models. Hence, the EIH-LM model's ability to predict the loudness functions from threshold to 50 dB SPL is worth noting.

(2) Above 50 dB SPL, there are two problems with the predicted loudness function. Firstly, the model loudness grows at a slower rate than human loudness for the corresponding frequencies. Secondly, it saturates at high sound pressure levels, approximately 70 dB SPL, and the saturation threshold decreases with increasing frequency.

(3) Below 500 Hz, the EIH-LM model failed completely.

Loudness Level Function and Iso-Loudness Contours

The predicted loudness level function for 500 Hz matches the human loudness level function well from 10 to 90 dB SPL. For 2 kHz, the model matches well up to about 70 dB SPL. For 3 kHz, the model function does not match that for human, but matches well with the human function for an 8-kHz tone. In the model, 4 kHz is the upper frequency limit, hence the relationship of 3 kHz tone to the upper limit of 4 kHz can be likened to that of 8 kHz to the upper frequency limit of the human audition range. For the model, the number of neurons excited by 3-kHz tone is relatively less than that by tones from 500 Hz through 1 kHz. This seems to indicate that the reason that the loudness growth rate of a tone close to the upper frequency limit is slower than that for a tone toward the mid-frequency range is the physical limitation.

Controversy Concerning Loudness Function

Comparing the prediction results for loudness functions and for loudness level functions, the conclusions drawn from the two results seem to contradict each other. The predictions for the loudness function indicate that the EIH-LM model falls far short of the observed data while those for the loudness level functions and the iso-loudness contours indicate that the model performs much better. Which result should one believe? To answer this question, one has to consider what these functions represent and how they are obtained. A loudness function specifies the growth of the absolute loudness of a tone as a function of SPL and was obtained by the magnitude production, magnitude estimation, doubling, and halving. These methods depend on people's innate ability to associate number systems with loudness perception. The loudness level functions and the iso-loudness contours were obtained by the loudness matching experiment and represent the loudness growth of a tone relative to that of the reference, a 1-kHz tone. Therefore, there is no conflict in the information conveyed by the two types of prediction.

The question is whether one believes the validity of the loudness function. At the present time, there are controversial views concerning the sone scale. Some firmly believe in its validity, while others question it. There is no doubt that everyone has a sense of the relative magnitude of the numbers in the real number system and also of the loudness of a tone at various SPL. However, the metaphor of the number system over loudness judgements is a very subjective process. The data obtained vary widely from subject to subject and from institution to institution, and are laden with biases.

On the other hand, the loudness function for a 1-kHz pure tone derived by Fletcher and Munson used results from a loudness matching task and was based on the loudness

summation of binaural vs. monaural listening and on the loudness summation of two tones far apart in frequency. The function was further verified based on the study of the neuron population across the basilar membrane and the firing rate of the neurons. The loudness level functions obtained by them seem believable, if one believes that loudness perception is proportional to the total neurons excited or to the total firing rate in the brain. Consequently the loudness function predicted by this type of model (i.e. EIH-LM) is believable and the comparison of the predicted loudness function to the loudness function obtained by Fletcher and Munson is valid. The loudness function for 1-kHz tone obtained by Fletcher and Munson was compared to that presented by Scharf in Fig. 3.3, and the two functions are similar.

Detection Thresholds

The predicted detection thresholds for pure tones were determined mainly by the thresholds of the lowest level detector, which in turn were determined by the envelope shape of the cochlear filters. Therefore, the detection thresholds of pure tones were determined by the middle ear transfer function, as reported by Rosowski (1991). The predicted detection thresholds by EIH-LM4, EIH-LM5, and EIH-LM6 for five complex stimuli showed that although the mean thresholds of the stimuli observed in human subjects were consistently higher than those predicted, the relative differences between detection thresholds of the complex stimuli and that of the 1-kHz pure tone were very similar between the predicted and observed stimuli. For example, if the observed detection threshold for a square wave was much higher than that for 1-kHz tone, the same was true for the predicted thresholds of square wave and pure tone. The ability of the EIH-LM model to predict correctly the

detection thresholds of the complex stimuli strengthens the view of loudness summation.

8.3.2 Effects of Various Parameters on Predictions

Effect of Temporal Integration

Effect of Rule 2 The 40-msec maximum integration time (rule 2) was the main cause of the mis-prediction by the EIH-LM model for tones below 500 Hz. The removal of the 40-msec limit in the EIH-LM2 model improved the model predictions for tones below 500 Hz significantly. For example, the model-loudness function for 250 Hz matched that for human below 55 dB SPL, like the predicted loudness functions above 500 Hz. The predicted loudness level function for 250 Hz matched that of human less well, i.e. the predicted and the observed functions matched only up to less than 40 dB SPL. This indicates the predicted loudness growth rate of 250 Hz relative to that of 1 kHz still falls short.

This improvement suggested that temporal loudness summation probably depends on a fixed number of cycles of the pure tone rather than a fixed amount of time elapsed. As it was found (Sivian and White, 1933) that it took about 3 cycles to establish the pitch of a tone regardless of the frequency of the tone, perhaps it is also true that it takes a fixed number of cycles to establish loudness perception. This result also suggests that uniform functional units across the whole BM are more likely than fixed elapsed time.

Effect of Rule 3 The third temporal integration rule had no effect on pure tones, but it had strong effect on complex stimuli. This rule tends to reduce the number of intervals used for integration because it favors low-frequency components rather than the high-frequency components. Whenever there is a relative smaller interval following several

large intervals, the rule is likely to be activated causing the EIH model to stop searching for more intervals. The magnitude of effect depended on the stimulus type. This rule seems to be responsible for the suppression of background noise and hence good performance for speech recognition tasks for the case of speech in noise. This effect was somewhat erratic and has no specific physiological or psychophysical observations to support it. From the loudness prediction point of view, it appears to be a poor choice.

Loudness Definition over Time

The main effect of the two definitions of loudness used in the EIH-LM model was on speech because of its large variations of sound intensity over time. Zwicker, in his loudness prediction for speech over time, had used the 95th percentile point of the calculated loudness over time to represent the overall loudness of a sentence. Fastl (1977) has shown that a speech-like noise was equally loud to running speech when the noise level was set to the maximum sound level of the speech in dB (A). In the loudness matching experiment conducted in this thesis, the loudness of a sentence was equally loud to a 1-kHz pure tone whose sound pressure level was equal to that of the sentence at the 90th percentile point of the instantaneous sound pressure levels. All this evidence suggests that the overall loudness of speech is determined by its peak loudness. This modification does not affect the conclusions for pure tones, because the mean of the predicted loudness of a pure tone is nearly the same as the maximum.

Saturation of Predicted Loudness at High Level

The saturation problem observed in the predicted loudness functions for pure tones at high SPL was also observed with complex stimuli. The parameters of the EIH-LM model

effective in changing the predicted loudness growth rate at high SPL included (1) the cochlear filter set; (2) the threshold range of the level detectors in each channel and across channels; and (3) spacings between the threshold levels of the detectors.

Effect of Cochlear Filters on Predicted Loudness

The extension of the cochlear filter set from 7.2 to 12 kHz in EIH-LM3 moved the saturation threshold level higher and increased the slopes of the predicted loudness functions at high SPL. Further, it balanced the loudness growth rates of other frequency tones relative to that of the 1-kHz tone, particularly those of the high-frequency tones at mid to high SPL and those of the low-frequency tones at low to mid SPL. Hence it improved the predicted iso-loudness contours and made them more like the observed ones. Further extension of cochlear filters to include higher frequencies would continue to improve the saturation problem for the predicted loudness functions and to balance the predicted loudness growths between frequency components. The highest characteristic frequency of the cochlear filter in EIH-LM3 model was 12 kHz, while the highest frequency perceivable is believed to be 20 kHz in human and 60 kHz in cat. The improvement could be substantial if all filters were included. The difficulty would be the availability of neural tuning curves for generating the needed cochlear filter set.

This improvement implies that the spreading of the excitation pattern along the basilar membrane upward and downward to include more channels with the increasing sound pressure level of a pure tone is a main contributor to the loudness growth of the tone. However, at high sound pressure levels, i.e. 70 to 80 dB SPL, it seems the whole basilar membrane would be excited. In order for the loudness to continue growing, as it apparently

does, other physiological mechanisms would have to play a role. Also, counting of the channels alone would not provide the rate of loudness growth as demonstrated by observed loudness functions. For comparisons of the loudness growth of a tone to that of 1-kHz, the basilar membrane excitation pattern would be sufficient to provide agreement with human loudness level functions over a large frequency range except at low frequencies, such as below 250 Hz.

Effect of Level Detectors on Predicted Loudness

Threshold Range of Level Detectors Two types of modifications were made to test the effect of the overall threshold range of the level detectors in each channel: one was the range comparison between all channels; and the other was the range within each channel. For the first type of modification, the threshold range between the lowest and the highest level detectors in each channel was set to be the same for all channels, i.e. 60 dB, instead of varying with the CF of the filter. This modification increased the saturation threshold slightly toward higher levels. This modification was supported by the physiological observation that the threshold range of auditory fibers with similar CF in healthy cat ears was similar for frequencies below 10 kHz. The second type of modification, that increased the threshold range from 60 dB to 80 dB, improved the saturation problem for pure tones at a very high SPL. Its effect on complex stimuli was more pronounced. However, this range was not realistic comparing to the threshold range of auditory nerve, except one might view it as a crude estimation for rate-level functions .

Number of Level Detectors When the sensitivity threshold range of the level detectors remained the same and the spacings between the thresholds were equal on a

logarithmic scale, i.e. equal in dB/step, the number of level detectors per channel did not seem to have any effect on the predicted loudness as demonstrated by EIH-LM7 and EIH-LM8. The two models were identical except EIH-LM8 had more level detectors than EIH-LM7, and the loudness functions for the four stimuli predicted by the two models were the same.

Spacings between Thresholds of Level Detectors Within a channel, the spacing between the thresholds of the level detectors appeared to have a substantial influence on the slopes of the predicted loudness functions. Three types of spacing were examined: equal spacing on a logarithmic scale, equal spacing on a linear scale, and compressive spacing on a logarithmic scale. The equal spacing on a linear scale, if converted to a logarithmic scale, showed increasing density with increasing SPL, hence the thresholds mainly clustered at high end of range. Variable spacing on the logarithmic scale arranged the thresholds so that the thresholds clustered at the low end of the range. Thus the three types actually represented the extremes of a continuum of possible threshold arrangements. The slopes of the predicted loudness functions corresponding to these three types of arrangement, clustered at high end, equal spacing, and clustered at low end, were steep, medium, and shallow, as shown in predictions of EIH-LM9, EIH-LM7, and EIH-LM12.

The loudness functions predicted by EIH-LM7 matched better than those predicted observed loudness functions, by EIH-LM9 and EIH-LM12. The threshold distribution of the level detectors in EIH-LM11 slightly clustered toward high end of the threshold range, much less so than in EIH-LM9. The loudness function predicted by it was a marked improvement over EIH-LM7, indicating that, by manipulation of the spacings between level detectors, it was possible to achieve a good match between predicted and observed loudness

functions. However, based on the neural data in cat concerning the distribution of the neurons according to their sensitivity thresholds, the configuration of the level detectors in EIH-LM12 was closer to the observed distribution of neuron thresholds, hence its predicted loudness function was more believable. Furthermore, if the threshold distribution of the auditory nerves, as observed by Liberman, were to be simulated by the level detectors, the characteristic of the loudness functions predicted by EIH-LM12 would be even stronger, i.e. extremely rapid growth of loudness above threshold, very low saturation threshold, and no growth in the saturation zone. The implication of this scenario was rather clear, i.e. loudness coding by the total number of the neurons excited was inadequate. On the other hand, one could view the configuration of the level detectors in EIH-LM7 and EIH-LM11 as a crude composite form of a rate-level function, and the improvement of the loudness functions predicted by these models seemed to indicate the necessity of rate-level functions to loudness prediction and that the total firing rate would be a more likely loudness coding scheme.

8.3.3 Complex Stimuli

(1) Complex sound at low sound pressure levels: This model's loudness prediction has much faster loudness growth than the human data. It seems some kind of suppressive mechanism is lacking in the model to provide the mutual suppression of the loudness of individual components.

(2) Complex sound at high sound pressure levels: The model falls short again. This seems to be the same problem as with the pure tone at high sound pressure levels. An increase in the number of threshold levels in the model seems to improve the situation. This fact seems to indicate that the rate level information and the threshold distribution are important for

the loudness growth.

(3) Synchrony is not used for loudness predictions in the cases presented in this dissertation and appears not to play a role in the total loudness. However, it may play an important role in partial loudness. This needs further investigation.

(4) The psychophysical tests used for evaluation of the physiologically-based auditory model are not sensitive enough to pinpoint the problems of the model. These tests can only evaluate the overall performance of the model.

8.4 Limitations of the EIH-LM Model

The EIH model used as the basis of the loudness model, EIH-LM, was by no means complete in its composition of physiological components and its simulation of those components. A brief discussion of the shortcomings of the model's approximations of these components in terms of the physiological observations in animals and in terms of those in human is presented as follows:

Cat Physiology

(1) The EIH model, in using a set of linear cochlear filters, ignored the level-dependent compression effect observed in the displacement functions of basilar membrane at CF.

(2) Any nonlinear effect on the middle ear transfer function due to the neural efferent loop, such as the acoustic reflex at high sound pressure level, was also ignored.

(3) The input-output transduction function of the inner hair cell was approximated by a half-wave rectifier which is a nonlinear operation. However, the compression seen in the input-output function that reduces the output dynamic range to about 30 dB was ignored

in the model.

(4) Since the function of the outer hair cell is not well-understood, there was no explicit representation of the outer hair cell function in the model.

(5) Two aspects of the statistical representations of the neural firings on the auditory nerves are presented in the EIH model, i.e. the thresholds of the fibers and the periodicity of the firing due to the phase-locking property of the fibers. However, the rate-level functions, another important aspect of the firings, were greatly simplified to have only two states, activated or not activated.

(6) Several approximations of the auditory nerves by the five level detectors in each channel seemed inaccurate. Firstly, the tapering of the threshold range of the five level detectors from 60 dB at low frequency to 20 dB at high frequency disagrees with the observed cat neural data. As observed by Liberman (1978), the range of the sensitivity thresholds of fibers with similar characteristic frequency (CF) is rather constant across frequencies below 10 kHz. Secondly, the thresholds of five level detectors are equally spaced logarithmically over the threshold range implying that fibers of similar characteristic frequency are distributed equally around the five levels. This approximation also disagrees with observations from cat neural data. Based on the extensive study by Liberman (1978) on the spontaneous rates and the threshold sensitivity of the primary auditory fibers in cats raised in sound chambers, the fibers could be divided into three groups: low-, mid-, and high-thresholds. The percentages of fibers for the three groups were 61%, 23%, and 16% respectively. The homogeneity of these fibers was such that The thresholds of the low- and mid-threshold groups was relatively homogeneous, ranging from 10 to 15 dB, comparing to the thresholds of the high-threshold group, spanning over 50 dB. Furthermore, the total

threshold range rarely reaches 60 dB.

(7) Another important nonlinear phenomenon, two-tone suppression, observed in the animal data was also not simulated by the model. The lack of this phenomenon is perhaps not crucial when only the periodicity of the stimulus, e.g. the pitch of speech, is of the interest, but is of particular importance when masking effects are needed in prediction of loudness of a stimulus.

(8) Spontaneous activities and Poisson character of neurons firings were not considered by the model either. If rate information were included in the model, then these aspects of firings might be important to the loudness prediction.

Human Physiology

When a model based on the animal physiology is used to predict human psychophysical data, it is necessary to bear in mind the difference between human and animal physiology. Although it has been shown that anatomical and physiological functions among the vertebrates are very similar, the interpretation of the prediction results still needs to be done with caution, due to the lack of human physiological data.

(1) The characteristic frequencies (CF) of the cochlear filters in the original model ranged from about 180 Hz to 7.2 kHz. In cat, this range covers a segment of basilar membrane from 3.9% to 58.2% of the total distance from the apex, while in human, the range corresponds to the segment from 15.5% to 78.7%. The later set of cochlear filters was extended to include filters with CF up to about 12.5 kHz, i.e. 69.1% for cat and 89.7% for human. In the case of humans, the lack of cochlear filters at the low frequency region left out an important area where loudness growth is quite unique and very interesting. To

remedy the problem, a proposal of obtaining a set of cochlear filters to simulate human cochlear filters and to cover the low frequency region has been presented in Section 8.5.

(2) Another concern was the bandwidths of the cochlear filters. It has been suggested that the human cochlear filters are more sharply tuned than those of the cats. Although the human neural tuning curves are not available, a method that may reveal the bandwidths of human cochlear filters to some degree using distortion product measurements has been suggested by Allen (1991). The data collected by Kimberley (1991) using this method does indeed shows that the bandwidths of human cochlear filters are narrower than those of the cats at frequencies below 10 kHz. Furthermore, the frequency dependency of the bandwidth on the characteristic frequency in the high frequency region observed for cat is not observed in human.

(3) The middle ear acoustic reflex in human as discussed by Møller (1983) occurs as low as 65 dB SPL. Due to the mis-prediction of the loudness at high sound pressure levels by the EIH-LM model, the lack of the middle-ear acoustic reflex is immaterial unless the model is improved to be able to predict loudness at high SPL.

(4) Among the three temporal integration rules used in the EIH model, the 40-msec maximum limit on the total elapsed time (rule 2) and the variable maximum limit varying from interval to interval (rule 3) have no apparent support from the physiological data. Rule 1 that limits the total integration time to a fixed 20 intervals is supported to some degree by the psychophysical data concerning pitch perception, detection thresholds, and temporal loudness summation. The 20-interval rule causes the temporal integration time to be frequency dependent as has been observed by some workers, although the exact integration times do not agree as the integration data do not agree across experiments by

different workers.

(5) The total firing rate as the composite of the firing rates of neurons with similar characteristic frequency and different sensitivity thresholds has been believed by many to be the explanation for the large intensity dynamic range observed in the human audition (Fletcher and Munson, 1933, 1937, 1938, 1940, 1950; Wever, 1949; Whitfield, 1978; Viemeister, 1988). In the EIH-LM model, since the rate-level functions of the auditory nerve fibers were not simulated, the neural coding correlated to loudness was defined to be the total number of neurons excited and the number of times they are excited over the defined integration time across all channels.

The effects of these limitations of the EIH model on loudness predictions have been detailed in the previous section. In general, these limitations caused the EIH-LM model not able to predict the full dynamic range as observed in human audition and not able to provide proper masking between neighboring frequency regions.

8.5 Recommendations for Further Research

As discussed in Section 8.4, there are many limitations in the EIH model in terms of its approximation of the functions in auditory physiology. If the EIH model is to be used to estimate the functions of human auditory system, then many areas of the EIH model should be improved. However, depending on the context of predictions, certain modifications are more relevant than others.

8.5.1 Cochlear Filters

In order to provide a set of cochlear filters that simulates human cochlear more closely, the following procedure is proposed: Based on the assumption that the functions of basilar

membrane are the same for human and for cat when the places on the basilar membrane are expressed in percentage distance from the apex as shown in Fig. 2.1, then the cat cochlear filters with CF ranging from 185 Hz to 12.5 kHz could be transformed to the approximated human cochlear filters with CF ranging from 34 Hz to 4500 Hz. The procedure is as follows:

1. Extract the cat middle-ear transfer function from the cat cochlear filters. It has been demonstrated that the middle ear transmission function reduces the low-frequency gain and increases the mid-frequency gain. It has also been demonstrated that the middle ear transfer function is the main cause of elevated thresholds in the low-frequency region (Rosowski, 1991).
2. Convert each cat filter, free of the middle ear influence, from the frequency domain to the place domain. This is done by identifying each filter in terms of its percent distance from helicotrema on the basilar membrane using the cat parameters for the cochlear frequency-position map (Eq. 2.1) described in Chapter 2.
3. Convert each filter back to the frequency domain using the human cochlear frequency-position map. By this procedure the converted filters are estimates of human cochlear filters. Note that these cochlear filters have CF's shifted toward lower frequencies.
4. Lastly, incorporate the human middle ear transmission characteristics into the estimated human cochlear filters.

The new set of approximation for human cochlear filters should have bandwidths narrower than the original cat filters. In order to improve the accuracy of the bandwidth estimation of human cochlear filters, a method proposed by Allen (1991) using distortion product measurement could be useful.

8.5.2 Level Detectors

It has been demonstrated that the distribution of level detectors has a strong effect on loudness predictions. However, the importance of accurate approximation of auditory nerves by level detectors is not limited to the loudness prediction, but applies to other applications also, such as speech recognition. Therefore, the representation of the level detectors should reflect the distribution of the auditory nerves closely.

8.5.3 Two-tone Suppression

As shown in the review of the loudness calculation methods developed since 1920's, the ability to take masking effects into account was essential for accurate prediction of loudness of complex sounds. It was also shown how difficult it was to estimate the mutual masking effect between neighboring frequency components within a critical band and between critical bands. The problem is further complicated by the level-dependency of the masking effect. The loudness predictions for complex stimuli in this dissertation also demonstrated the failure of the EIH-LM model for lack of a stronger mechanism to handle the mutual masking effect.

The suggestion for improving the EIH-LM model for loudness prediction is to incorporate the two-tone suppression model developed by Delgutte (1989, 1990) into the EIH-LM model to provide masking effects, hence to provide level-dependent cochlear filters. Delgutte had demonstrated that two-tone suppression play a major role in masking and has presented a model to simulate two-tone suppression. The model was able to predict the masking pattern for pure tones and many other stimuli.

8.5.4 Rate-level Function

The rate-level function should be incorporated. It has been shown that rate-level functions for neurons of different sensitivity are slightly different in their slopes and hence their dynamic ranges. Furthermore, also shown was the upward shifting of rate-level functions for the high-threshold neurons in the presence of noise (Young and Barta, 1986). This latter observation may be an explanation of the ability of the loudness of a narrow-band noise to continue to grow in gated noise that eliminates the loudness contribution from the spread of excitation.

One way to incorporate various rate-level functions is to replace the half-wave rectifier and the level detectors in the EIH model by an inner hair cell model, such as the one by Allen (1983), which provides the means and variances of the rates for corresponding stimulus levels.

8.6 Conclusion

The EIH-LM model was found to be reasonably accurate in predicting loudness for pure tones at low to moderate sensation levels. Predictions for complex stimuli were not as good, but the results were nevertheless revealing with respect to the physiological factors affecting loudness summation.

The factors affecting loudness summation for pure tones were found to be fewer and simpler than those affecting loudness summation for complex stimuli. Firstly, loudness prediction results for pure tones show that the range of frequencies covered by the set of cochlear filters is an important factor, because loudness summation depends on spread

of excitation along the basilar membrane. Secondly, the prediction results show that the number of neurons excited in each channel contribute to the loudness summation. Thirdly, in the model the lack of rate information for the excited neurons causes the predicted loudness function to saturate and to have a small dynamic range, indicating that neural firing rate is also an important factor for loudness summation. These three factors, spread of excitation, number of neurons responding, and firing rate of response, are necessary to account for loudness summation of complex stimuli. However, additional factors are required because of mutual masking between neighboring components of the stimulus. This mutual masking would reduce the excitation contributed by each channel. Thus, the loudness function for a complex stimulus would be steeper than for a pure tones because of spread of excitation, but mutual masking would reduce the undesirably steep loudness function predicted by the EIH-LM model in its present form. Mutual masking might correspond to physiologically-observed two-tone suppression phenomenon. However, two-tone suppression has been shown to be more prominent at high frequencies and for high threshold neurons. Hence, it is not clear how much mutual suppression at low intensity could be accounted for by physiological two-tone suppression. This aspect of the model should be studied further. Another factor affecting loudness predictions of the model is integration time. Predictions for pure tones by the EIH-LM model with and without a fixed integration time showed that integration using a fixed number of cycles gave much better results than integration using a fixed integration time. It may appear that this result conflicts with data on observed firing rate and stimulus duration. However, one should remember that the EIH model is based on excitations (positive level crossings) at several threshold levels and not on firing rate. Nevertheless, the effect and interpretation

of temporal integration in this model deserve further investigation.

Appendix A

Loudness Matching Experiment

A.1 Subjects

A total of 16 people were tested for hearing thresholds. Out of 16, 12 subjects were able to participate in the Loudness Matching experiment. Some identifying information about each subject relevant to the experiment is summarized in Table A.1. The left-most column contains the subject's identification number (ID). The fourth column contains the subject's hearing threshold in dB SPL for 1-kHz tone in the more sensitive ear, identified in the right-most column.

A.1.1 Subject Audiograms

The detection thresholds for pure tone between 100 and 8000 Hz were measured for both ears of each subject. Von Békésy's tracking method was used. The procedure started with a 1000-Hz tone and moved first toward 100 Hz, then returned to 1000 Hz and moved toward 8000 Hz. The measurement used the same equipment setting as the monaural loudness matching experiment, in which the tones were delivered to the subjects' ear canals via ER-2 tubephone (Etymotic Research) and a disposable foam eartip.

Subj ID.	Sex	Age (years)	Hearing Threshold (1 kHz, dB SPL)	Ear Tested
1	F	37	7.0	L
3	M	47	4.0	R
4	M	31	6.0	R
5	M	31	9.0	R
6	F	30	6.0	R
7	F	32	14.0	L
8	M	23	9.0	L
9	F	34	5.0	R
13	M	30	11.0	R
14	M	27	8.0	L
15	M	21	8.0	L
16	M	20	9.0	L

Table A.1: The first column show each subject's identification number (ID). The next two columns provide information on the subjects (Sex and Age). The remaining two columns specify the subject's more sensitive ear, the one used for the Loudness Matching experiment, and the hearing threshold for a 1-kHz pure tone in dB SPL for the more sensitive ear.

The more sensitive ear of the two was used for the monaural loudness matching experiment. Twelve audiograms of the more sensitive ears are shown in Fig. A.1, one for each subject. In each plot, the solid curve was derived from the arithmetical means of every pair of adjacent peak and valley values of the tracking data. The open circles represent the averaged normal pure tone detection thresholds using ER-2 earphone (Etymotic, 1987). The frequencies of the tones are 125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz.

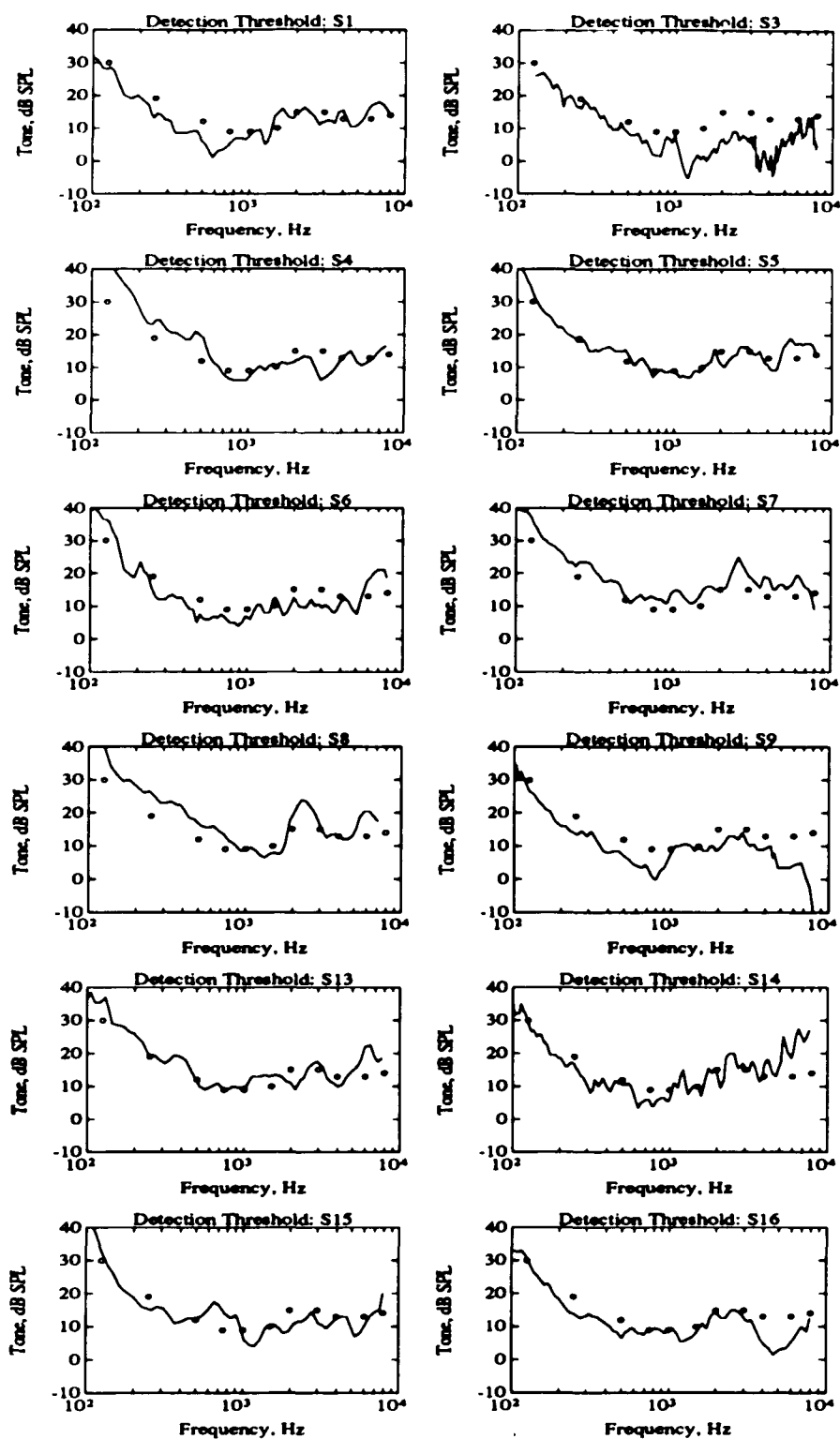


Figure A.1: Pure Tone Detection Thresholds for Each Subject in dB SPL Between 100 and 8000 Hz. The open circles represent the averaged normal using the same earphone transducer set.

A.2 Test Protocols

As mentioned in Chapter 6, there were six test sessions in the loudness matching experiment: three experiment conditions tested by two adjustment methods. The experiment was designed to be repeated measures, hence all 12 subjects had to be tested for all six test sessions. In order to avoid the systematic carrying-over effect of one test session on another throughout the experiment, the six test sessions were arranged according to the six sequences in the Latin-square design shown in Table A.2. A latin square of 6 elements is a matrix of six rows by six columns, where the permutation in each row and in each column is unique. Furthermore, in each row or each column every number has to appear once, i.e. no two numbers in a row or column are the same. The six columns correspond to the six test sessions, and the six rows are the six unique permutation sequences. There were 12 subjects tested in the experiment, hence each sequence was used to test two subjects. The subjects tested for each sequence are listed in the two right-most columns.

Test Protocol Using Latin Square								
Seq. ID.	Test Session Numbers						Subjects	
1	1	2	3	4	5	6	S4	S14
2	2	6	4	3	1	5	S5	S13
3	3	4	5	6	2	1	S9	S16
4	4	1	6	5	3	2	S6	S7
5	5	3	1	2	6	4	S3	S8
6	6	5	2	1	4	3	S1	S15

Table A.2: Six test sessions were arranged according to the Latin-square design to obtain six unique sequences. Since there were 12 subjects, each sequence was used to test two subjects. Left-most column (column 1) is the sequence identification number, columns 2 through 7 are the six test session numbers, and columns 8 and 9 show the subjects tested using the sequence.

A.3 Loudness Matching Experiment Instructions

Following are the written instructions for each of the six test sessions, two sessions for each experimental condition. The two sessions correspond to the two methods of adjustment. In session A, the reference stimulus is adjusted in level to match the loudness of the test stimulus. In session B, the test stimulus is adjusted in level to match the loudness of the reference stimulus. For condition Q, the tasks in two sessions are the same, the difference is with the stimuli to be adjusted.

A.3.1 Instructions for Test Session 1: Session A of Quiet Condition (Q)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus will be one of three signals: a narrow band noise, a speech sentence, or a pulsive sound. The second stimulus of the pair will be either a pure tone or a narrow band noise. Your task is to adjust the intensity of the second stimulus so that the second stimulus sounds **EQUALLY LOUD** to the first stimulus.

Match the loudness using the **BRACKETING METHOD**:

1. Turn the knob up (clockwise) until you can clearly tell that the second stimulus is louder than the first stimulus.
2. Turn the knob down (counterclockwise) until the second stimulus is softer than the first stimulus.
3. Then turn the knob up again until the second stimulus sounds **EQUALLY LOUD** to the first. Repeat step(s) 2 or 2 and 1 if necessary.
4. When the equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.3.2 Instructions for Test Session 2: Session A of Partial-loudness Condition (P)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus contains a wideband noise and one of three signals, a narrow band noise, a speech sentence, or a pulsive sound. The second stimulus of the pair will be either a pure tone or a narrow band noise. Your task is to adjust the intensity of the second stimulus so that the second stimulus sounds equally loud to the **LOUDNESS** of the signal alone in the first stimulus. The loudness of a signal in the presence of another sound will be referred to as **PARTIAL LOUDNESS** of the signal later.

Match the loudness using the **BRACKETING METHOD**:

1. Turn the knob up (clockwise) until you can clearly tell that the second stimulus is louder than the PARTIAL LOUDNESS of the signal in the first stimulus.
2. Turn the knob down (counterclockwise) until the second stimulus is softer than the PARTIAL LOUDNESS of the signal in the first stimulus.
3. Then turn the knob up again until the second stimulus sounds equally loud to the PARTIAL LOUDNESS of the signal in the first stimulus. Repeat step(s) 2 or 2 and 1 if necessary.
4. When the equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.3.3 Instructions for Test Session 3: Session A of Overall-loudness Condition (O)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus contains a wideband noise and one of three signals: a narrow band noise, a speech sentence, or a pulsive sound. The second stimulus of the pair will be either a pure tone or a narrow band noise. Your task is to adjust the intensity of the second stimulus so that it sounds equally loud to the TOTAL LOUDNESS of the first stimulus, i.e. the wide band noise and the signal combined.

Match the loudness using the BRACKETING METHOD:

1. Turn the knob up (clockwise) until you can clearly tell that the second stimulus is louder than the TOTAL LOUDNESS of the first stimulus.
2. Turn the knob down (counterclockwise) until the second stimulus is softer than the TOTAL LOUDNESS of the first stimulus.
3. Then turn the knob up again until the second stimulus sounds equally loud to the TOTAL LOUDNESS of the first. Repeat step(s) 2 or 2 and 1 if necessary.
4. When equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.3.4 Instructions for Test Session 4: Session B of Quiet Condition (Q)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus of the pair will be either a pure tone or a narrow band noise. The second stimulus will be one of three signals: a narrow band noise, a speech sentence, or a pulsive sound. Your task is to adjust the intensity of the second stimulus so that the second stimulus sounds EQUALLY LOUD to the first stimulus.

Match the loudness using the BRACKETING METHOD:

1. Turn the knob up (clockwise) until you can clearly tell that the second stimulus is louder than the first stimulus.

2. Turn the knob down (counterclockwise) until the second stimulus is softer than the first stimulus.
3. Then turn the knob up again until the second stimulus sounds EQUALLY LOUD to the first. Repeat step(s) 2 or 2 and 1 if necessary.
4. When the equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.3.5 Instructions for Test Session 5: Session B of Partial-loudness Condition (P)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus of the pair will be either a pure tone or a narrow band noise. The second stimulus contains a wideband noise and one of three signals, a narrow band noise, a speech sentence, or a pulsive sound. Your task is to adjust the intensity of the signal in the second stimulus so that the LOUDNESS of the signal alone in the second stimulus sounds equally loud to the first stimulus. The loudness of a signal in the presence of another sound will be referred to as PARTIAL LOUDNESS of the signal later.

Match the loudness using the BRACKETING METHOD:

1. Turn the knob up (clockwise) until you can clearly tell that the PARTIAL LOUDNESS of the signal in the second stimulus is louder than the first stimulus.
2. Turn the knob down (counterclockwise) until the PARTIAL LOUDNESS of the signal in the second stimulus is softer than the first stimulus.
3. Then turn the knob up again until the PARTIAL LOUDNESS of the signal in the second stimulus sounds equally loud to the first. Repeat step(s) 2 or 2 and 1 if necessary.
4. When the equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.3.6 Instructions for test session 6: session B of Overall-loudness Condition (O)

There are 54 trials in this test session. In Each trial, you will hear a pair of stimuli repeated continuously. Each stimulus is about 2.4 seconds long. The first stimulus of the pair will be either a pure tone or a narrow band noise. The second stimulus contains a wideband noise and one of three signals, a narrow band noise, a speech sentence, or a pulsive sound. Your task is to adjust the intensity of the signal in the second stimulus so that the TOTAL LOUDNESS of the second stimulus, i.e. the wide band noise and the signal combined, sounds equally loud to the first stimulus.

Match the loudness using the BRACKETING METHOD:

1. Turn the knob up (clockwise) until you can clearly tell that the TOTAL LOUDNESS of the second stimulus is louder than the first stimulus.

2. Turn the knob down (counterclockwise) until the TOTAL LOUDNESS of the second stimulus is softer than the first stimulus.
3. Then turn the knob up again until the TOTAL LOUDNESS of the second stimulus sounds equally loud to the first. Repeat step(s) 2 or 2 and 1 if necessary.
4. When equal loudness is achieved, push the green button to stop the stimuli.
5. To go on to the next trial, turn the knob down (counterclockwise).

At the end of the experiment, the computer will solicit your comment about the experiment. A brief comment is appreciated. Then the experiment will terminate.

A.4 Thresholds of Six Test Stimuli

The purpose of this measurement was to examine the extent to which the detection thresholds of the stimuli used in the Loudness Matching Experiment differed from that of the 1 kHz pure tone. These data were obtained because the stimuli were presented to each subject at sensation levels relative to the subject's detection threshold for 1-kHz pure tone. Table A.3 contains the detection thresholds of the six stimuli for five subjects. The 50% detection thresholds were obtained using the method of adjustment (bracketing approach), the same method and the same system as used for the Loudness Matching Experiment. Each entry in the table is the average of three replications. A repeated measures (over subjects) showed that the square-wave and 1 kHz pure tone thresholds differed significantly ($p < 0.01$).

Stimulus thresholds (dB SPL)						
Subj. ID.	Speech	NBN-150	Square Wave	Pure tone	NBN-120	MPN
1.1	14.0	11.0	26.3	15.3	10.3	17.3
1.2	16.4	14.3	24.7	17.0	13.7	18.0
3	8.2	7.5	21.3	7.6	8.4	12.7
5	14.3	11.6	17.7	12.1	11.8	16.7
7	21.0	25.7	33.3	20.0	25.3	26.3
13	22.8	16.5	30.7	17.7	17.2	23.7
Mean	16.1	14.4	25.7	15.0	14.5	19.1
Std. Dev.	5.3	6.3	5.8	4.5	6.1	5.0
Std. Err.	2.2	2.6	2.4	1.8	2.5	2.0

Table A.3: The six stimuli are speech, NBN-150, square wave, 1-kHz pure tone, NBN-120, and modified pink noise. The first two rows are the retests for the subject 1. The last three rows are the means, the standard deviations, and the standard errors of the six stimuli.

Note that the pure tone thresholds measured in this way are uniformly higher than the detection thresholds measured using Békésy's tracking method (see Table A.1). Both methods were implemented on the same system using the same equipment, transducer, and calibration method. In the tracking method, the detection thresholds were calculated by

averaging the peaks and the valleys of the Békésy's tracking. A comparison of the threshold data collected from the two methods showed that thresholds obtained with the adjustment method corresponded mainly to the peaks in the Békésy's tracking method. This results suggested that the difference between the thresholds obtained by the two methods was a result of differences in how the criterion used in calculating the threshold. For example, if the peaks of the Békésy's tracking were used to estimate the threshold, then both methods would have yielded the same results.

A.5 Loudness Matching Results

Tables A.4 to A.9 show the means and standard deviations for all experimental conditions averaged across subjects. All stimulus levels are specified in dB re the subjects' detection thresholds for 1-kHz pure tone, i.e. in dB SL1k.

A.5.1 Mean Response Levels for Quiet Condition

All twelve subjects completed both test sessions for the Quiet Condition, i.e. two adjustment method. The data in both Table A.4 and A.5 are for all 12 subjects.

Group Means & Standard Deviations in dB SL1k (Adjust REFERENCE Stimulus)							
Ref. Sig.	Test Sig.	30 dB SL1k		50 dB SL1k		70 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	31.1	6.3	51.0	5.6	66.2	3.4
	NBN150	32.2	3.8	47.2	4.2	65.8	3.6
	SQW	30.8	6.7	56.9	7.6	71.7	6.3
NBN120	Speech	35.0	6.7	54.6	5.5	69.3	4.2
	NBN150	32.0	2.6	50.4	1.6	69.4	0.9
	SQW	31.6	8.3	63.4	6.6	75.3	4.9

Table A.4: The means and standard deviations of the response levels obtained by adjusting the reference stimulus for test stimuli at 30, 50, and 70 dB SL1k. The data are for all 12 subjects who completed the experiment for this test session.

Group Means & Standard Deviations in dB SL1k (Adjust TEST Stimulus)							
Ref. Sig.	Test Sig.	30 dB SL1k		50 dB SL1k		70 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	33.1	6.0	49.7	6.8	66.7	7.3
	NBN150	34.4	3.4	54.1	3.4	71.0	1.8
	SQW	31.8	3.1	42.3	4.0	53.4	5.6
NBN120	Speech	31.4	4.1	47.4	6.2	63.4	8.6
	NBN150	31.7	2.5	50.3	1.1	69.4	1.3
	SQW	31.9	3.2	41.7	4.5	51.7	7.1

Table A.5: The means and standard deviations of the response levels obtained by adjusting the test stimulus for reference stimuli at 30, 50, and 70 dB SL1k. The data are for all 12 subjects who completed the experiment for this test session.

A.5.2 Mean Response Levels for Partial-loudness Condition

For both test sessions, i.e. two adjustment methods, Subject 6 was not able to complete the experiment. Therefore, the data are for 11 subjects.

Group Means & Standard Deviations in dB SL1k (Adjust REFERENCE Stimulus)							
Ref. Sig.	Test Sig.	50 dB SL1k		60 dB SL1k		70 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	36.0	9.2	49.5	7.5	63.0	7.5
	NBN150	34.2	7.8	48.5	5.8	60.0	5.2
	SQW	31.4	10.7	55.9	8.4	68.6	7.3
NBN120	Speech	37.6	7.9	53.0	6.7	64.3	5.9
	NBN150	34.7	6.0	50.9	5.8	63.0	3.8
	SQW	31.2	10.0	59.0	8.3	69.8	6.3

Table A.6: The means and standard deviations of the response levels obtained by adjusting the Reference stimulus for test stimuli partially masked by the background noise. The test stimuli were at 50, 60, and 70 dB SL1k, and the background noise was at 50 dB SL1k. The data are for 11 subjects who completed the experiment for this test session.

Group Means & Standard Deviations in dB SL1k (Adjust TEST Stimulus)							
Ref. Sig.	Test Sig.	30 dB SL1k		50 dB SL1k		70 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	51.7	4.2	59.6	4.8	73.2	5.4
	NBN150	50.8	4.2	62.7	3.7	75.6	3.3
	SQW	52.5	2.5	57.6	4.1	69.6	11.2
NBN120	Speech	50.7	4.7	59.7	5.6	71.5	6.3
	NBN150	50.7	2.9	61.1	2.8	73.7	2.7
	SQW	52.2	2.6	56.9	3.7	66.7	6.7

Table A.7: The means and standard deviations of the response levels obtained by adjusting the Test stimulus for reference stimuli at 30, 50, and 70 dB SL1k. The test stimulus was partially masked by the background noise at 50 dB SL1k. The data are for 11 subjects who completed the experiment for this test session.

A.5.3 Mean Response Levels for Overall-loudness Condition

The data in test session A are for all 12 subjects who completed the experiment. The data in test session B are for 9 subjects, S1, S3, S6, S8, S9, S13, S14, S15, and S16. The data for the other 3 subjects, S4, S5, and S7 were incomplete, hence their were not included.

Group Means & Standard Deviations in dB SL1k (Adjust REFERENCE Stimulus)							
Ref. Sig.	Test Sig.	40 dB SL1k		50 dB SL1k		60 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	61.5	8.4	59.6	5.8	64.8	6.6
	NBN150	61.4	8.7	61.8	8.1	63.6	5.8
	SQW	60.2	6.8	61.4	8.5	66.8	6.8
NBN120	Speech	63.9	9.4	64.2	7.8	67.0	6.2
	NBN150	63.6	9.0	62.2	7.1	65.2	6.2
	SQW	64.0	9.2	64.9	7.4	68.8	7.6

Table A.8: The means and standard deviations of the response levels obtained by adjusting the Reference stimulus for test stimuli plus the background noise. The test stimuli were at 40, 50, and 60 dB SL1k, and the background noise was at 50 dB SL1k. The data are for 12 subjects who completed the experiment for this test session.

Group Means & Standard Deviations in dB SL1k (Adjust TEST Stimulus)							
Ref. Sig.	Test Sig.	70 dB SL1k		72 dB SL1k		75 dB SL1k	
		means	Std.	means	Std.	means	Std.
PT1k	Speech	68.4	8.5	72.0	8.9	76.2	8.0
	NBN150	73.5	5.1	76.1	5.0	78.6	4.4
	SQW	65.7	7.7	66.4	7.5	69.6	8.0
NBN120	Speech	69.8	10.0	72.7	9.9	75.4	9.0
	NBN150	72.0	7.1	75.1	5.6	77.9	4.4
	SQW	63.5	7.8	67.6	9.1	68.5	8.6

Table A.9: The means and standard deviations of the response levels obtained by adjusting the Test stimulus so that the total loudness of test stimulus plus the background noise at 50 dB SL1k was matched to the loudness of the reference stimuli at 70, 72, and 75 dB SL1k. The data are for 9 subjects who completed the experiment for this test session.

A.6 Analysis of Variance for Experimental Variables

A second ANOVA was performed on the means of the three replications of the test conditions to examine the effects of the six experimental variables on the measured loudness levels. The six experimental variables examined were L (stimulus level), T (test stimulus type), R (reference stimulus type), A (adjustment method), C (experimental condition), and S (subject), (see Table 6.2 for details of the variables.) The results of the analysis are summarized in Table 6.4. In this section, the mean response levels for all the statistically significant variables and interactions are plotted and the detailed discussions are provided for some of the significant terms.

The analysis results showed that, of the six experimental variables, L, C, and S were statistically significant ($p \leq 0.001$). Among interactions common to all subjects, LC, AC, LTA, LTC, LAC, TAC, and LTAC were statistically significant ($p \leq 0.001$). Graphical plots for the main effects, L and C, and for this group of significant interactions are shown in Section A.6.1. Among interactions involving subjects, LS, TS, LTS, AS, LAS, TAS, LTAS, RS, RAS, CS, LCS, TCS, ACS, LACS, and TACS were found to be statistically significant ($p \leq 0.001$). Graphical representations for the main effect S and for this group of significant interactions are shown in Section A.6.2. Of all these significant effects, the largest ones were the main effect L, followed by interactions AS, LC, and the other two main effects C and S.

A.6.1 Significant Effects across All Subjects

The following are some observations from Figures A.2 through A.4 and from the available post-hoc analyses¹ (Tukey, 1949).

The mean responses for L, C, LC, and AC are shown in Fig. 6.9. The mean responses for the four three-way interactions, LAC, LTC, and LTA, are shown in Figures A.2 through A.4, and TAC is shown in Fig. 6.10. For purposes of illustration, the three-way interactions are plotted in two different ways. The plots corresponding to the first method of representation are shown in panels (a - c), and those corresponding to the second method in panels (d - f).

The post-hoc analysis showed that the mean responses to the three levels of L or the three conditions of C differed significantly. The variations between three levels were significantly large for both conditions Q and P. For condition O, the mean response levels to the low- and mid- levels were similar.

The significant interactions observed among L, T, A, and C can be attributed in part to asymmetries in the experimental conditions. For condition Q, for example, test stimulus and the reference stimulus were both in quiet, and consequently, the same set of levels was used for both the test and reference stimuli for the two methods of adjustment; i.e. when the reference stimulus was adjusted, the test stimulus was set at 30, 50, or 70 dB SL1k, and when the test stimulus was adjusted, the reference was also set at 30, 50, or 70 dB SL1k. In condition P, however, the reference stimulus in quiet was matched to the test stimulus in noise, resulting in an asymmetrical choice of stimulus levels. When the

¹The post-hoc analysis of significant interactions did not extend beyond two-way interactions because the more complex interactions involved too many contrasts for a reliable analysis.

reference stimulus was adjusted, the test stimulus was set at 50, 60, or 70 dB SL1k. When the test stimulus was adjusted, the reference stimulus was set at 30, 50, or 70 dB SL1k in order to cover the same range of interest. The situation for condition O was similar to that for condition P. As a consequence, the interactions involving A could be ascribed to these asymmetries. The post-hoc analysis of the AC interaction indicated that the mean responses obtained using the two methods were not significantly different for condition Q, but were significantly different for conditions P and O. Fig. A.2 shows that the differences between mean responses obtained by the two methods of adjustment for each experimental condition were also level dependent.

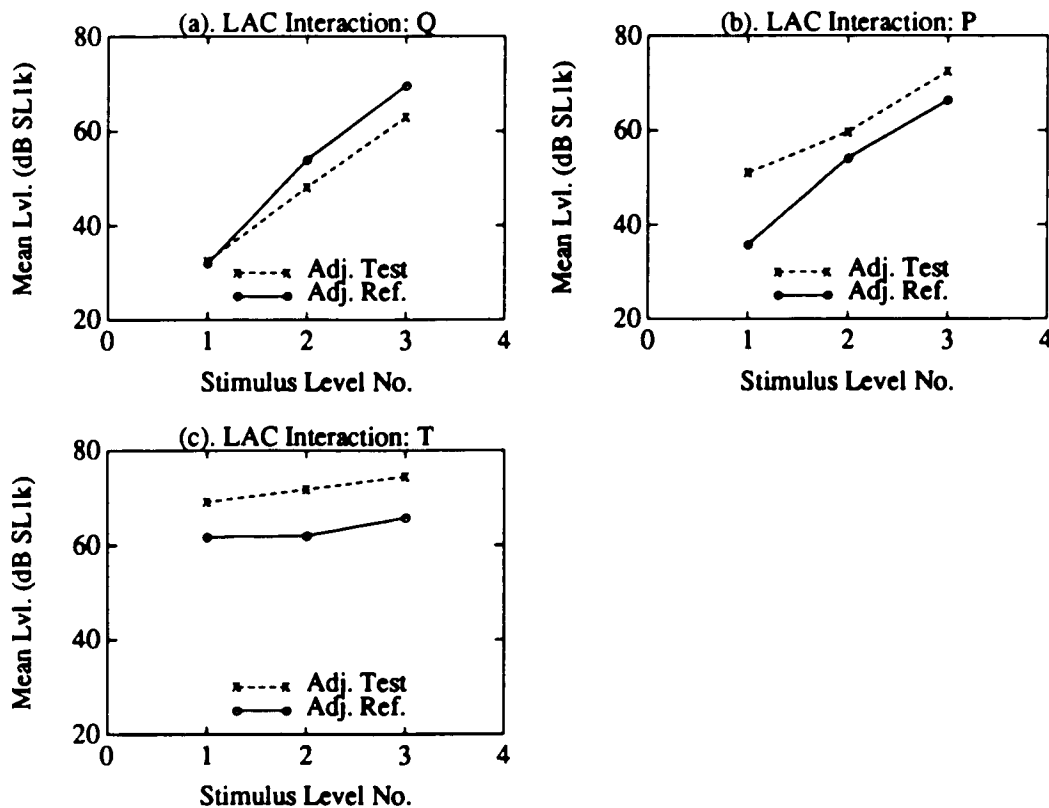


Figure A.2: Graphical Representation of LAC Interaction. Mean response levels averaged over variables T, R, and S are shown for the two adjustment methods (A) at three stimulus levels (L) of each condition (C). Panels (a) through (c) correspond to the three conditions Q, P, and O, respectively.

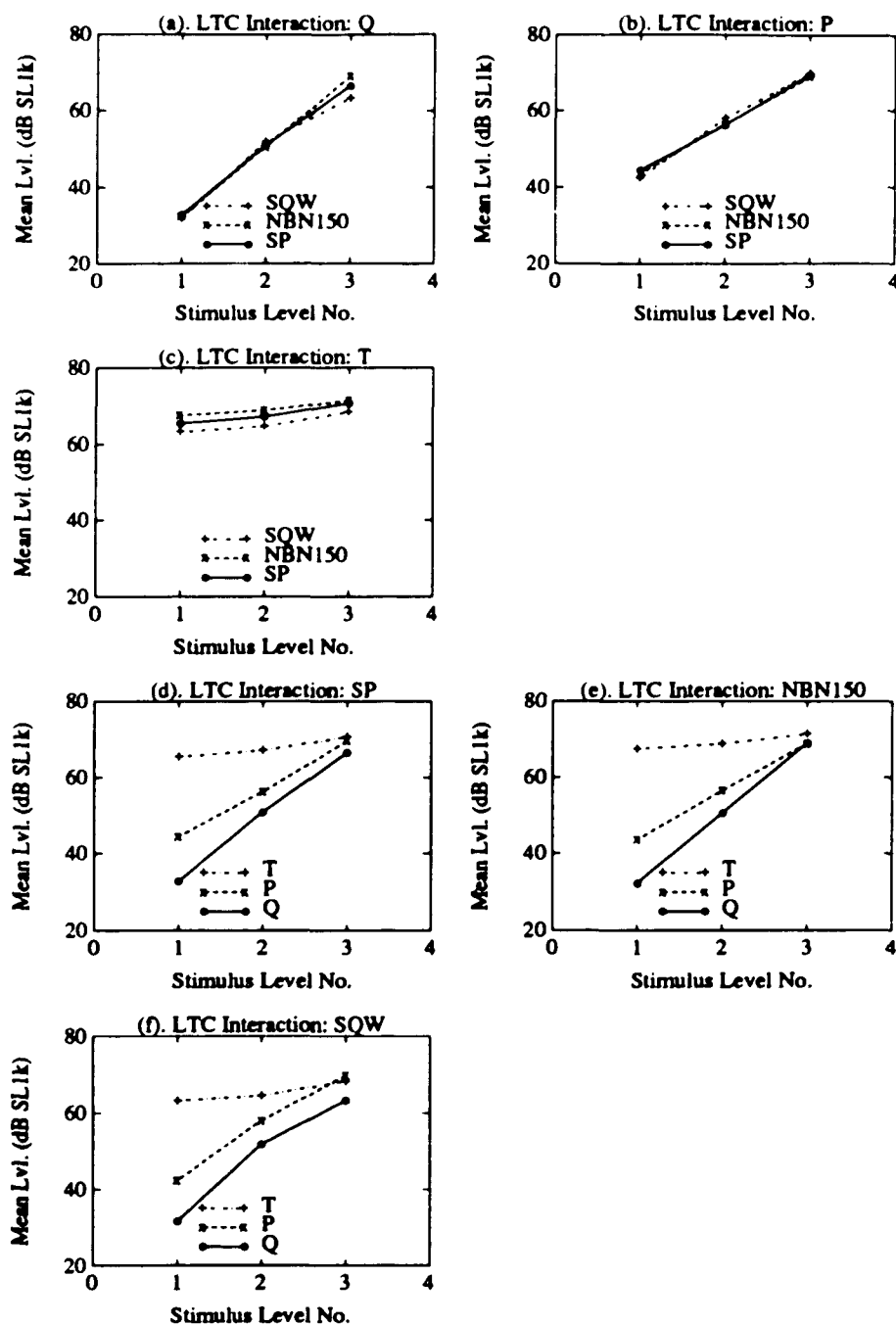


Figure A.3: Graphical Representation of LTC Interaction. Mean response levels for interaction LTC shown in two different ways. Panels (a - c) show the mean response levels for the three test stimulus types over stimulus levels for each condition with one condition for each panel. Panels (d - f) show the mean response levels for the three conditions at three levels for each test stimulus with one stimulus per panel.

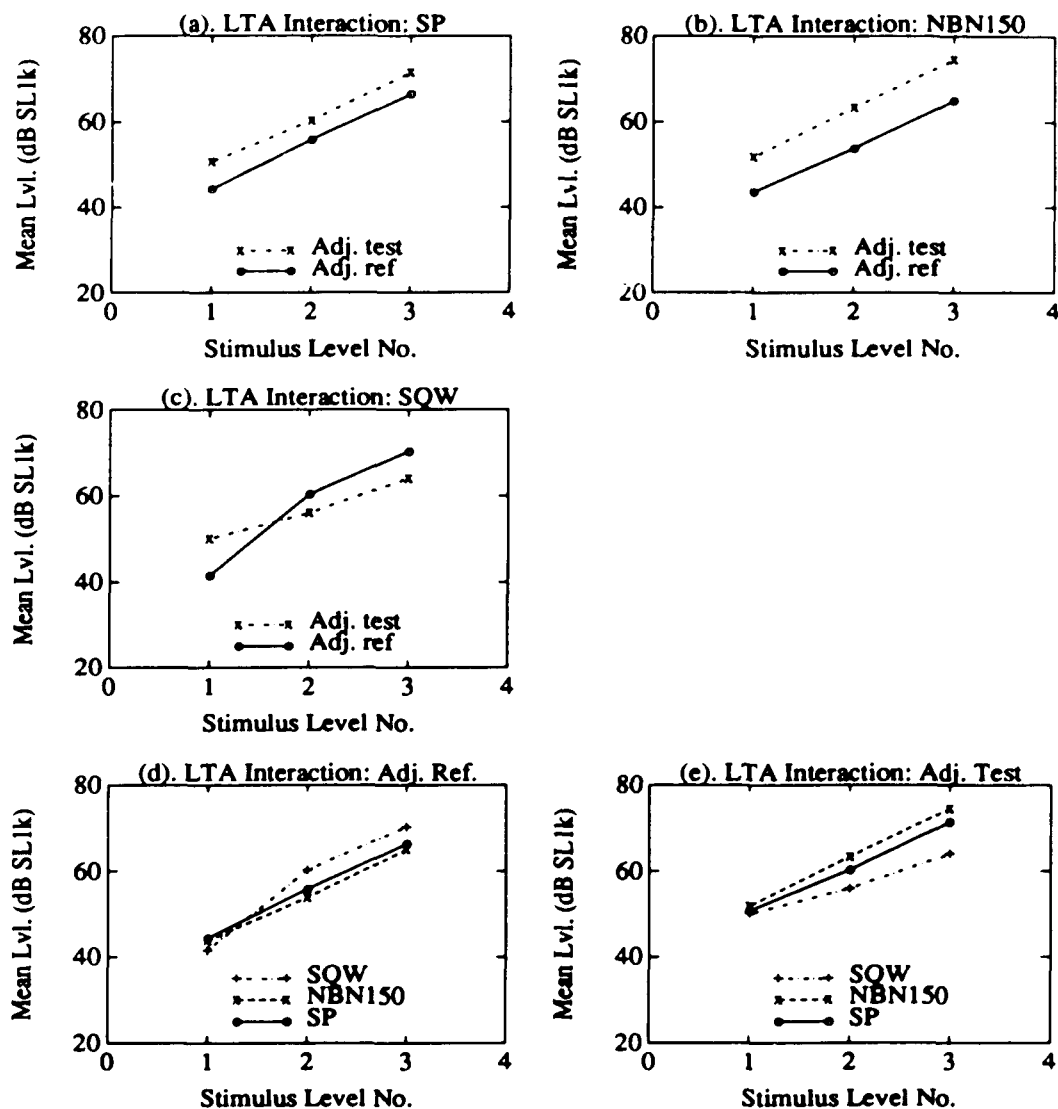


Figure A.4: Graphical Representation for LTA Interaction. Mean response levels for interaction LTA shown in two ways. In panels (a - c), mean response levels obtained using two adjustment methods at each stimulus level are plotted for three test stimuli in the three panels respectively. In panels (d - e), the mean response levels for 2-way interaction LT are plotted for the two adjustment methods in the two panels respectively.

Interaction LTAC

The effects of significant interactions common to all subjects as shown previously can be summarized in Fig. A.5 where nine plots for nine combinations of three test stimuli by three experimental conditions are shown. In each plot, the measured reference stimulus levels (open circles) are shown as a function of the test stimulus levels (fixed) and the measured test stimulus levels (asterisks) are shown as a function of the reference stimulus levels (fixed). Hence, the abscissa represents the test stimulus level in dB SL1k (sensation level re each subject's pure tone detection threshold at 1 kHz), and the ordinate represents the reference stimulus level in dB SL1k. When plotted in this way, the two curves connecting the two sets of symbols could be considered as loudness level functions corresponding to the two adjustment methods. As shown in the plots, the differences between the two functions are small for conditions Q and P and are very large for condition O. Hence, condition O is the main source of the variations for variable A, and the large differences between the mean responses to the two adjustment methods in condition P is artificial.

The biggest interaction LC is clearly illustrated in this figure. By comparing the loudness growth patterns across three test stimuli (T) for each condition, the pattern for SQW is seen to be different from those of SP and NBN150, especially at middle and high test stimulus levels, i.e. the loudness (reference) levels for SQW are higher than those for SP and NBN150.

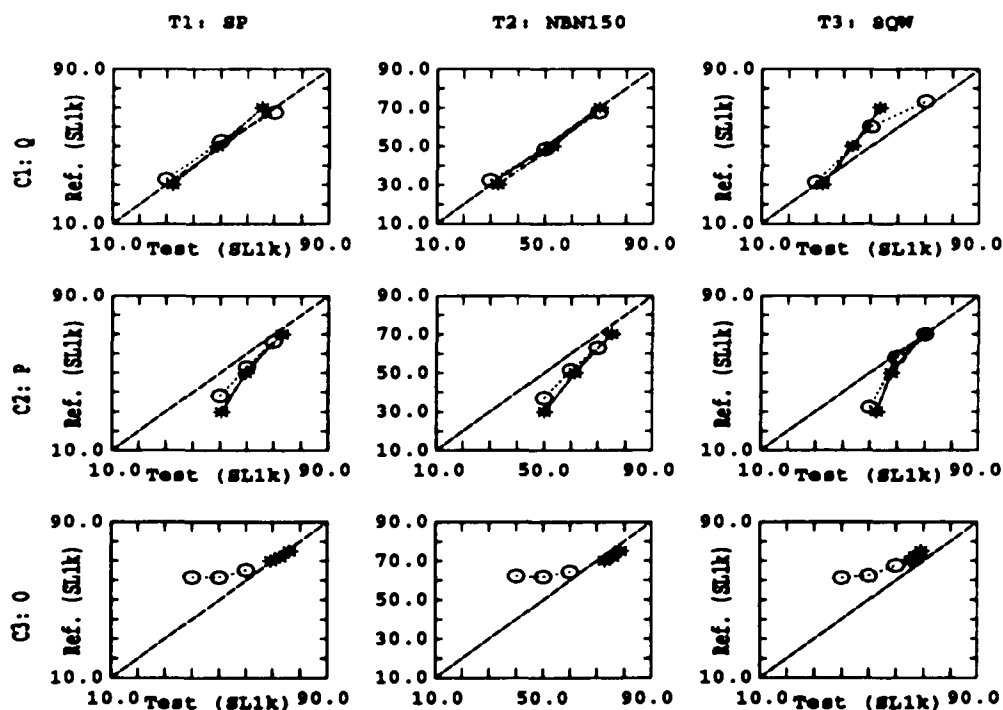


Figure A.5: Interactions Among the Experimental Variables, L (Stimulus Level), T (Test Stimulus), A (Adjustment Method), and C (Experimental Condition). Mean response levels averaged over the variables R (Reference Stimulus), and S (Subject) are shown. Nine plots for the nine combinations of three experimental conditions by three test stimulus types are arranged in three rows (conditions) and three columns (test stimulus). In each plot, open circles represent the mean loudness levels (reference levels) averaged across two reference stimulus types and eight subjects. Asterisks represent the mean test stimulus levels averaged in the same way.

A.6.2 Significant Effects Among Subjects

Another large main effect was the variations among subjects (S). The post-hoc analysis at $p < 0.01$ level showed that subjects could be subdivided into five groups: [S9], [S13], [S16, S15], [S3, S14], [S14, S8], and [S8, S1], ranked in ascending order of their overall mean levels, shown in panel (f) in Fig. 6.11. The variations between subjects were also found to have statistically significant interactions with A, C, T, L, and R, (listed in descending order of the magnitude of their interactions (F-ratio's)). The interactions between S and each of these five variables are shown in panels (a - e) of Fig. 6.11.

Significant Three-way Interactions

The significant three-way interactions involving subjects included LCS, TCS, RAS, TAS, LTS, LAS, and CAS. Among these interactions, those involving Adjustment Method variable, A, were much larger than the others. These interactions are shown in Figures A.6 through A.9. Each figure contains eight plots, one for each subject. Each plot shows the interactions between the remaining two main variables. In all four figures, subjects S8, S9, and S14 show distinctly different interaction patterns compared to the other subjects. Further, for these subjects the mean response levels obtained by adjusting the test stimulus were typically lower than those obtained by adjusting the reference stimulus whereas the reverse bias was shown by the other subjects. The magnitude of the differences between the mean response levels obtained by the two adjustment methods also varied widely between subjects, including differences as large as 30 dB (S15, CA interaction, Fig. A.6).

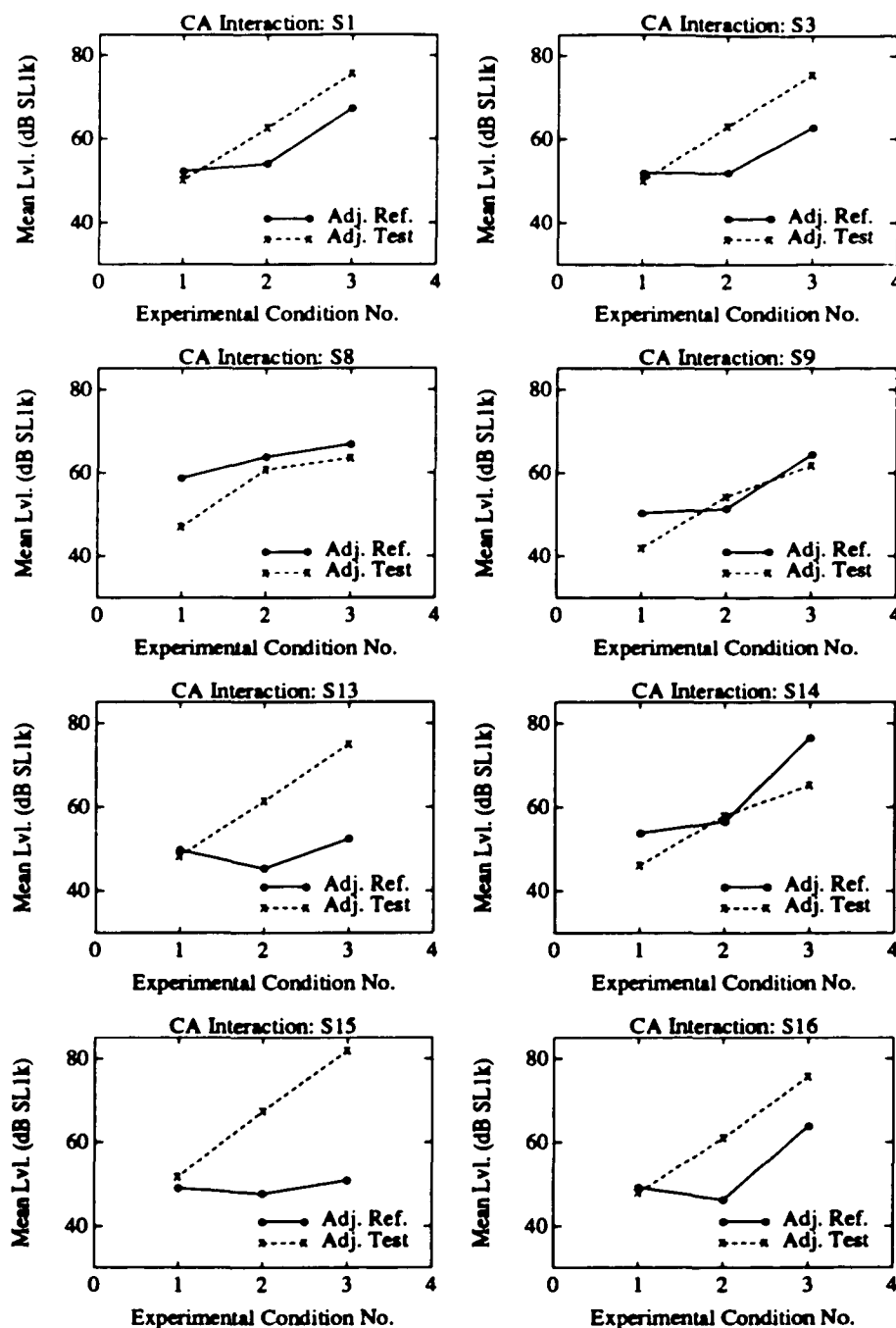


Figure A.6: Graphical Representation of CAS Interaction. Mean response levels obtained using the two adjustment methods (A) are shown for each experimental condition (C) and for each of the eight subjects. Each mean response level was obtained by averaging over three test stimulus levels (L), three test stimuli (T), and two reference stimuli (R). On the abscissa, experimental conditions 1, 2, and 3 correspond to condition Q (quiet), P (partial loudness), and O (overall loudness), respectively.

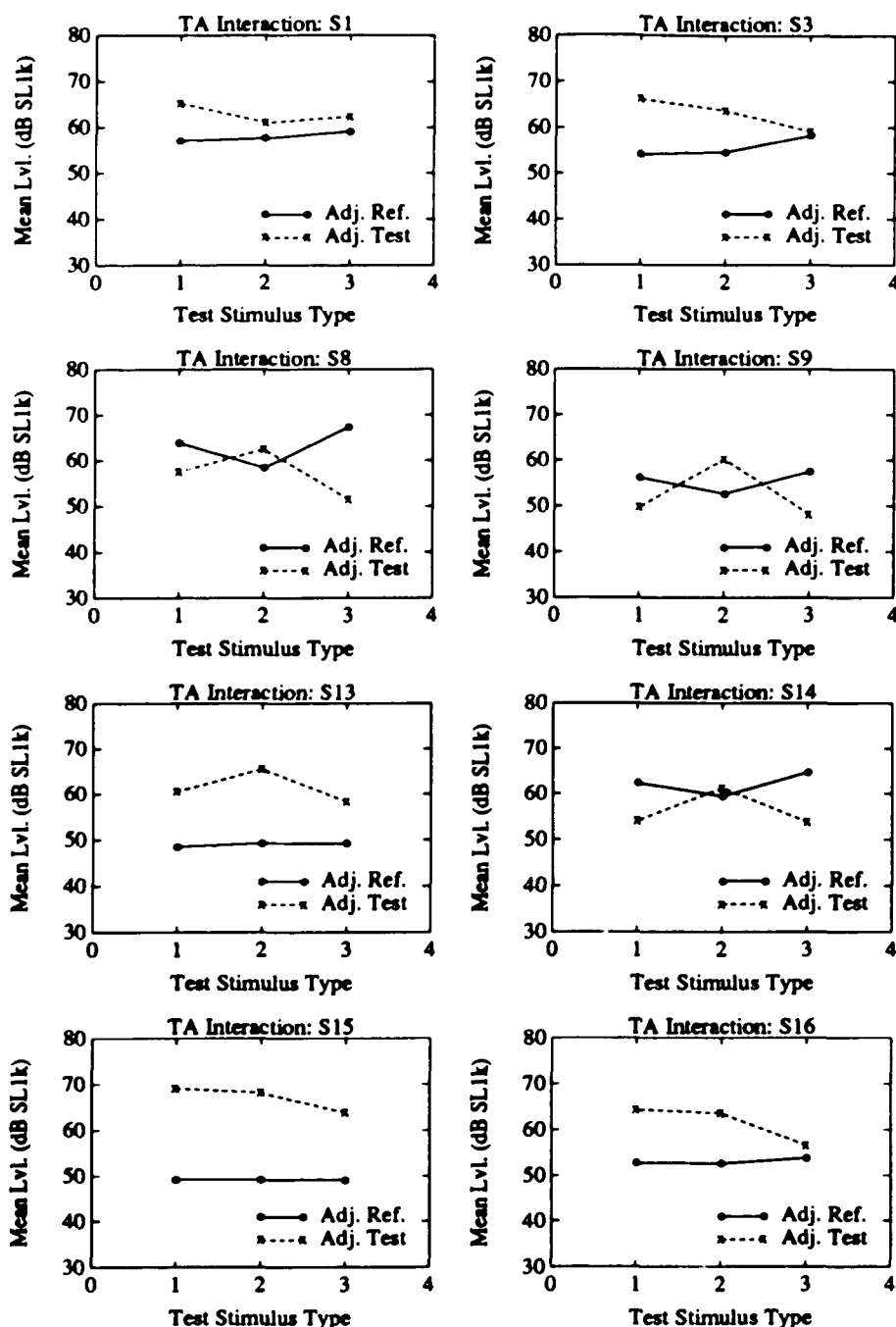


Figure A.7: Graphical Representation of TAS Interaction. Each diagram shows the interaction between Test Stimulus (T) and Adjustment Method (A) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over three test stimulus levels (L), three experimental conditions (C), and two reference stimuli (R). On the abscissa, test stimulus type 1, 2, and 3 correspond to speech (SP), narrow-band noise (NBN150), and square wave (SQW), respectively.

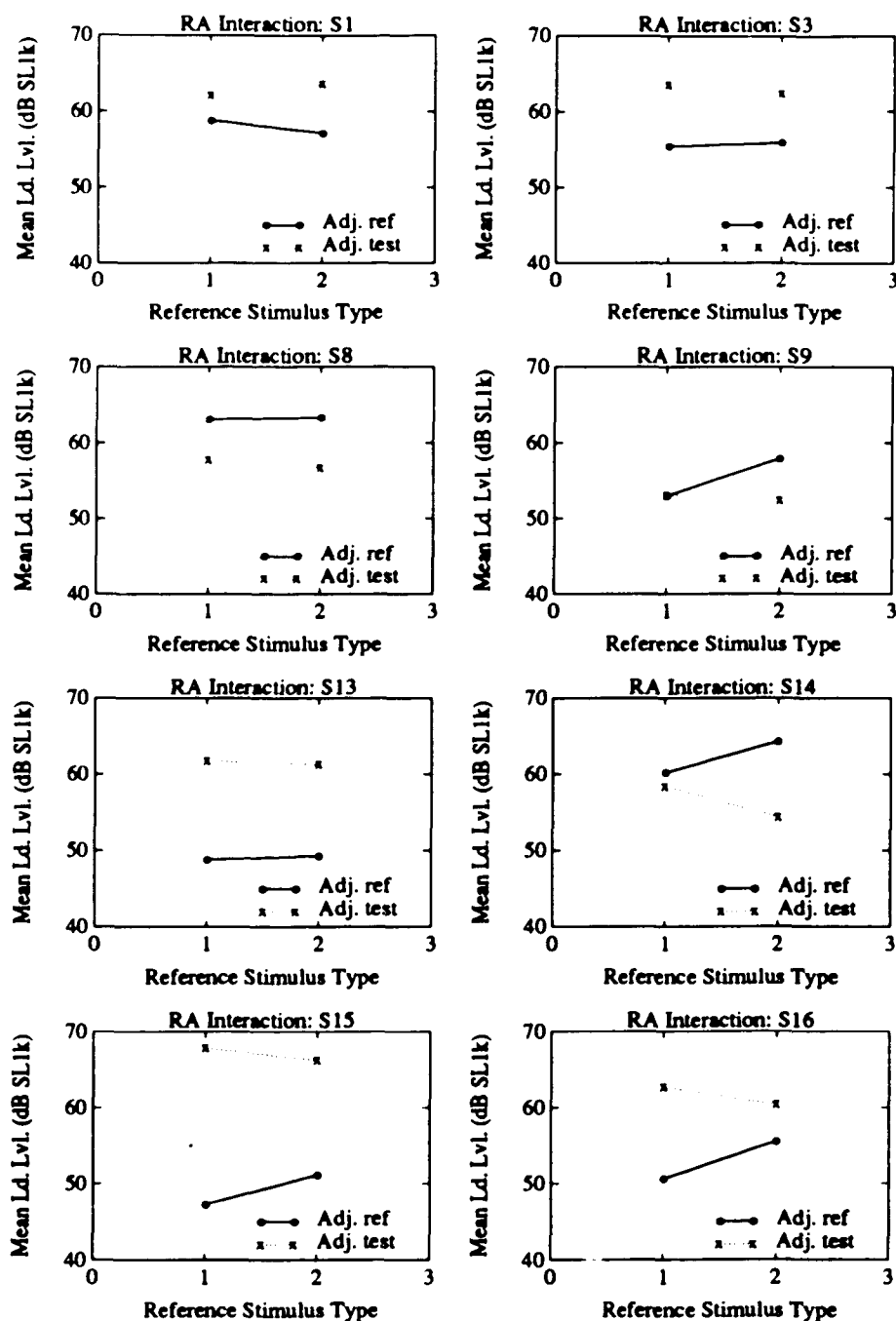


Figure A.8: Graphical Representation of RAS Interaction. Each diagram shows the interactions between the Reference Stimulus (R) and Adjustment Method (A) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over three test stimulus levels (L), three experimental conditions (C), and three test stimuli (T). On the abscissa, reference stimulus type 1, and 2 correspond to 1-kHz pure tone (PT) and narrow-band noise (NBN120), respectively.

The three 3-way interactions involving subjects, TCS, LCS, and LTS although statistically significant were relatively small. Graphical representations of these interaction are shown in Figures A.10 through A.12. In each figure, the interactions for TC, LC, or LT were similar across all subjects except S8, whose interaction patterns for experimental conditions P and O were different from other subjects in both TC and LC. For LTS interaction, S8 showed the least variation across the three test stimuli.

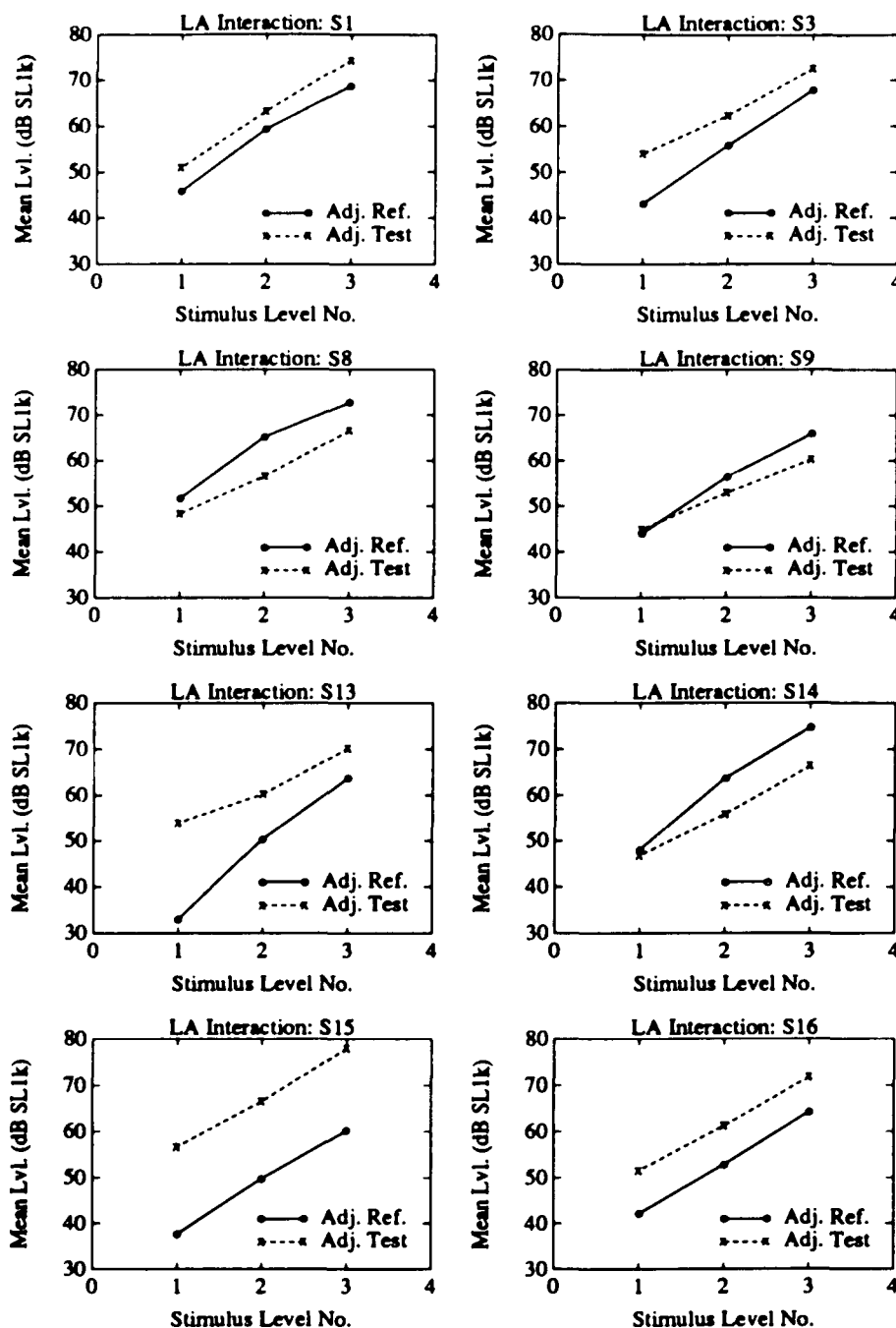


Figure A.9: Graphical Representation of LAS Interaction. Each diagram shows the interaction between Stimulus Level (L) and Adjustment Methods (A) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over two reference stimulus type (R), three experimental conditions (C), and three test stimuli (T). On the abscissa, stimulus level number 1, 2, and 3 correspond the low-, mid-, and high-level, respectively.

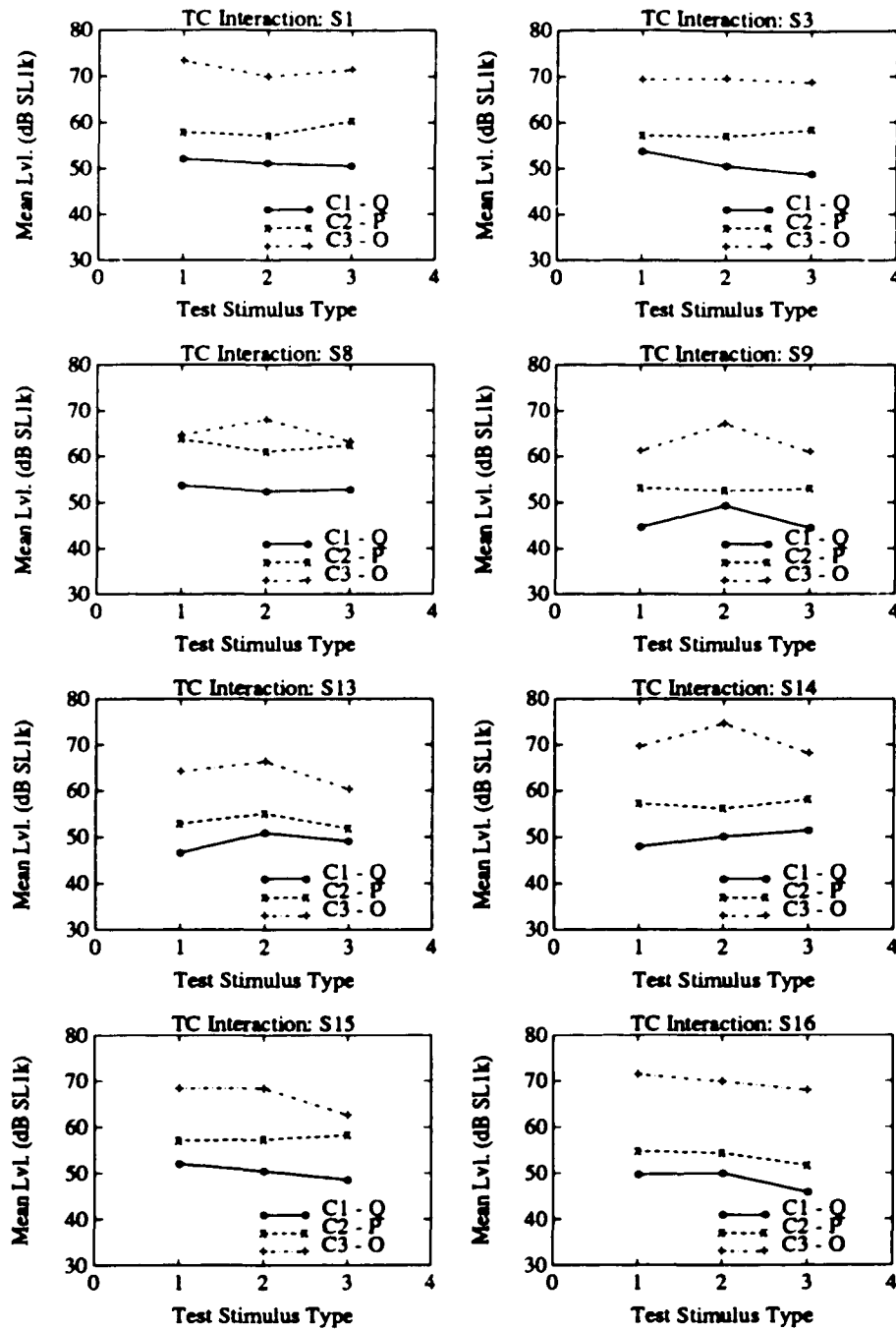


Figure A.10: Graphical Representation of TCS Interactions. Each diagram shows the interaction between Test Stimulus (T) and Experimental Condition (C) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over the reference stimulus type (R), the stimulus level (L), and the adjustment method (A). On the abscissa, the test stimulus type 1, 2, and 3 correspond SP, NBN150, and SQW, respectively.

Significant Four-way Interactions

Three significant 4-way interactions involving subjects were obtained, i.e. LTAS, LACS, and TACS. These interactions are shown graphically in Figures A.13 through A.18. In each figure, four rows of plots correspond to four subjects: the first figure of a set shows the interactions for subjects S1, S3, S8, and S9; while the second figure shows the interactions for subjects S13, S14, S15, and S16.

From Fig. A.13 and Fig. A.14, one can observe that subjects S1, S3, S13, S15, and S16 show similar interactions between Adjustment Method (A) and Experimental Condition (C). Although the magnitude of this interaction varies among subjects and across experimental conditions, the general pattern is the same. specifically, the mean response level obtained by adjusting the Test Stimulus (shown by the + symbols) is greater than that obtained by adjusting the Reference Stimulus for the Partial-loudness condition, and even greater for the Overall-loudness condition. For the Quiet condition, both methods of adjustment give essentially the same result. In contrast, subjects S8, S9, and S14 showed a different, more complex pattern of interactions. For the Quiet and Overall-loudness conditions, the mean response level obtained by adjusting the reference stimulus was higher than that obtained by adjusting the test stimulus, (a result opposite to that obtained by the other 5 subjects.) For the Partial-loudness condition, the data for these subjects were similar to the remaining subjects except that the difference between the mean loudness levels for the two methods of adjustment were much smaller.

Similar to the previous observation, graphical representations of the interaction LACS in Figures A.15 and A.16 show that subjects S1, S3, S13, S15, and S16 show similar interactions between Adjustment Method (A) and Stimulus Level (L). For these subjects,

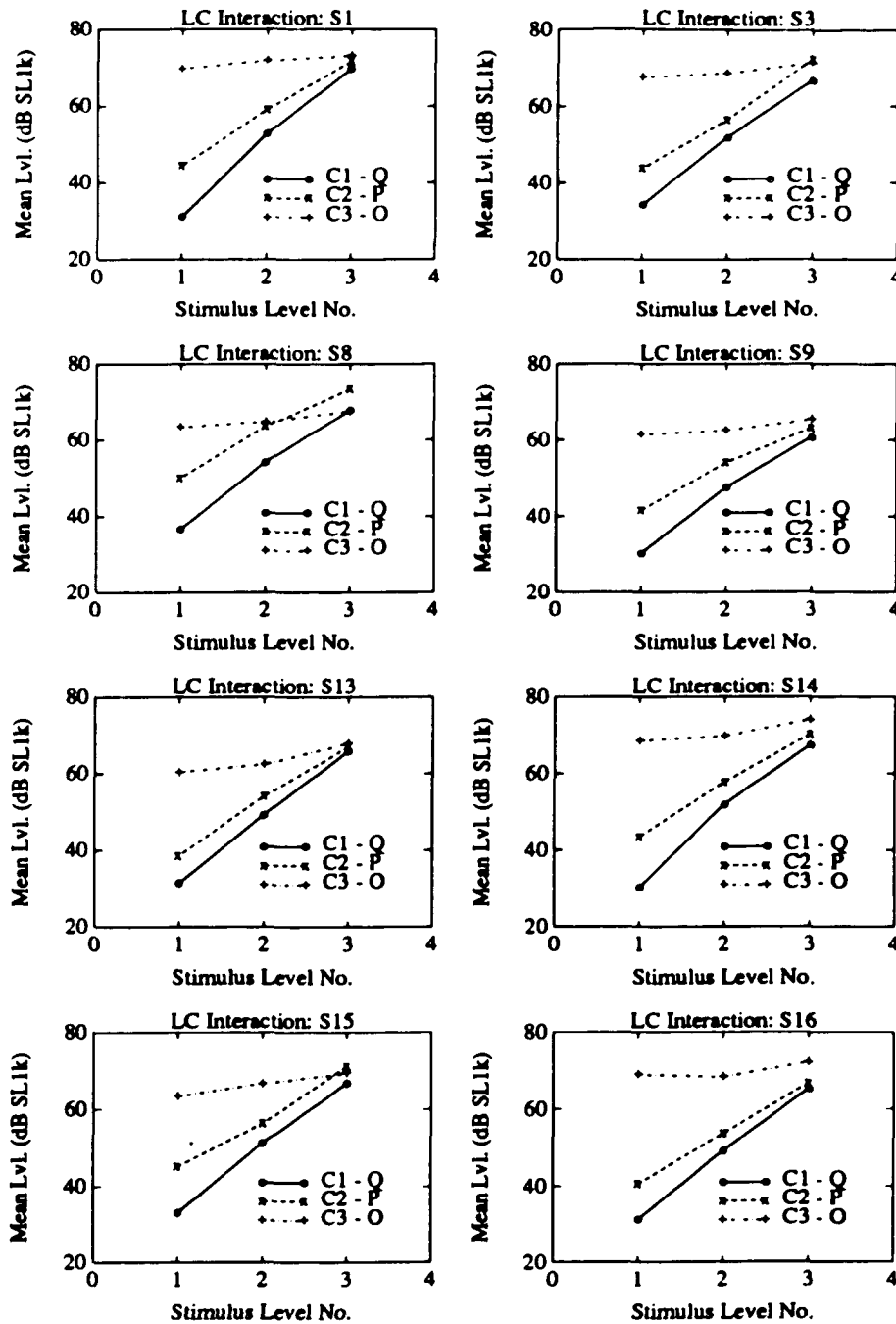


Figure A.11: Graphical Representation of LCS Interactions. Each diagram shows the interaction between Stimulus Level (L) and Experimental Condition (C) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over the reference stimulus type (R), the adjustment method (A), and three test stimuli (T). On the abscissa, stimulus level number 1, 2, and 3 correspond the low-, mid-, and high-level, respectively.

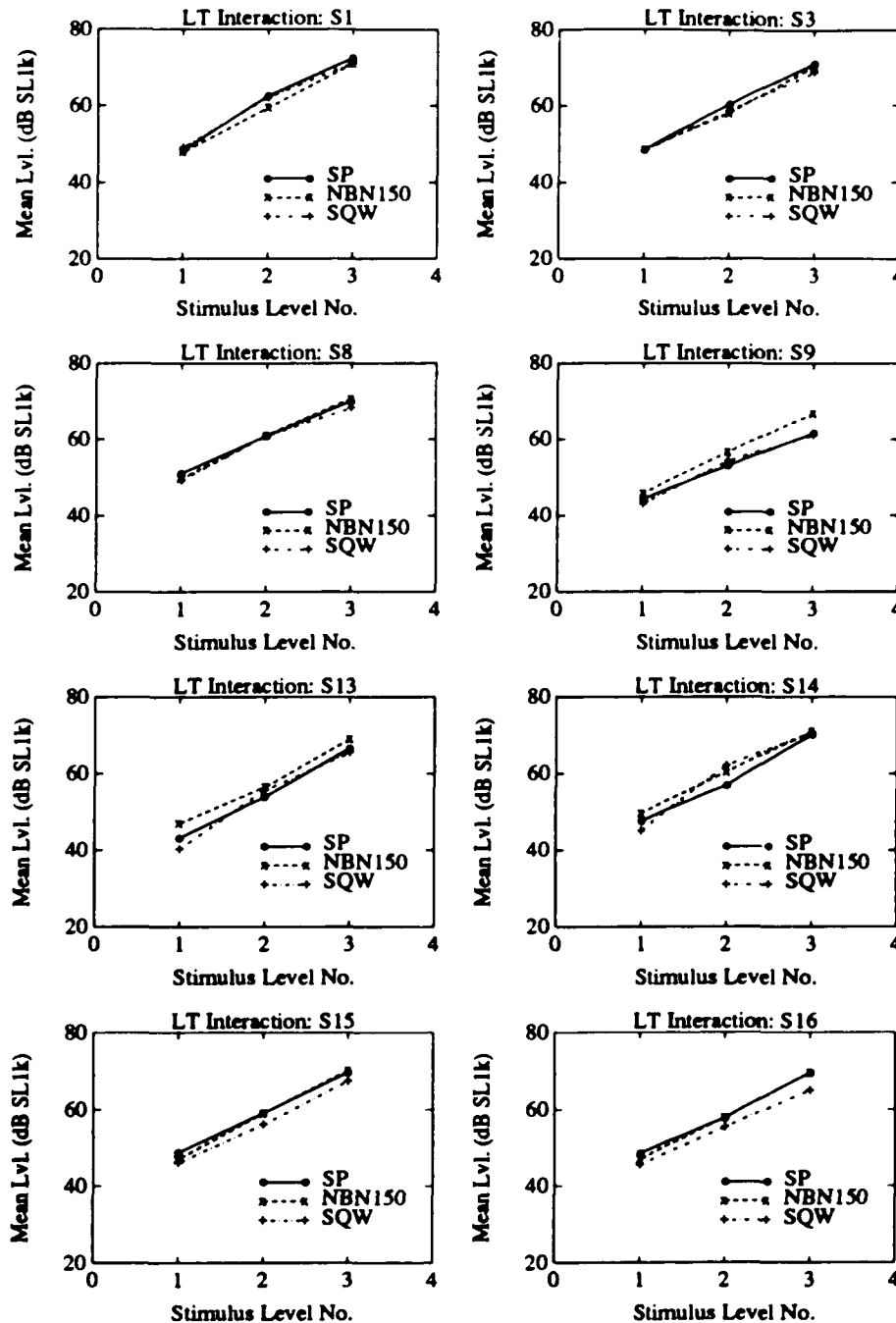


Figure A.12: Graphical Representation of LTS Interactions. Each diagram shows the interaction between Stimulus Level (L) and Test Stimulus (T) for each of the subjects. The ordinate in each diagram shows the mean response level obtained by averaging over two reference stimulus type (R), three experimental conditions (C), and two adjustment methods (A). On the abscissa, stimulus level number 1, 2, and 3 correspond the low-, mid-, and high-level, respectively.

the mean response levels obtained by adjusting the test stimulus were higher than those obtained by adjusting the reference stimulus in the Partial-loudness and Overall-loudness conditions, whereas the mean response obtained by the two methods were almost identical for the Quiet condition. The magnitude of the differences between the mean responses of the adjustment methods varies among subjects and decreases with levels in Partial-loudness condition. In contrast, the mean responses of subjects S8, S9, and S14 showed patterns different from those of the other 5 subjects. In Quiet condition, the mean response levels obtained by adjusting the reference stimulus were greater than those obtained by adjusting the test stimulus. For Partial-loudness and Overall-loudness conditions, the mean response levels obtained by the two methods were similar.

Graphical representations in Fig. A.17 and Fig. A.18 show interactions among Adjustment Method (A), Stimulus Level (L), and Test Stimulus (T) for each of the subjects. When the Test Stimulus is the narrow-band noise (NBN150), all the subjects show similar pattern in that the mean response levels obtained by adjusting the test stimulus were higher than those obtained by adjusting the reference stimulus, although the magnitude of the differences varied widely (up to 20 dB for S15.) When the Test Stimulus was speech (SP), subjects could be divided into two groups, similar to the previous observation, based on the patterns of the mean response levels. Subjects S1, S3, S13, S15, and S16 showed that the mean response levels obtained by adjusting the test stimulus were greater than those obtained by adjusting the reference stimulus across all stimulus levels. Subjects S8, S9, and S14 showed patterns of the mean response levels opposite to the other 5 subjects. When the Test Stimulus was square wave (SQW), Subjects S8, S9, and S14 showed very similar patterns of the mean response levels in that the mean response levels obtained by

adjusting the reference stimulus were greater than those obtained by adjusting the test stimulus at mid- and high-stimulus levels. For the other 5 subjects, the patterns are more complex. At low stimulus level, all five subjects showed the mean response levels obtained by adjusting the test stimulus were higher than those by adjusting the reference. At mid- and high-stimulus levels, all subjects, except S15, showed that the mean response levels obtained by the two adjustment methods were very close.

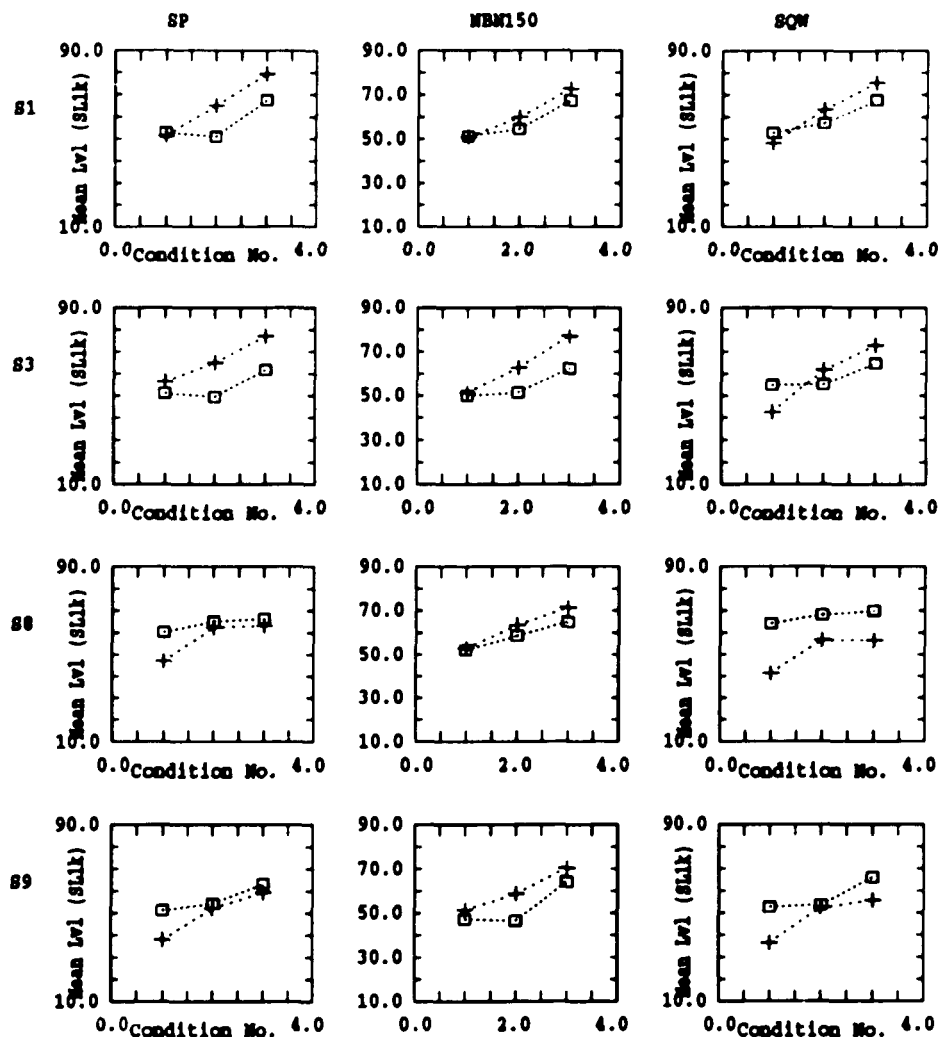


Figure A.13: Graphical Representation of TACS Interaction. Each diagram shows the interaction between Adjustment Method (A) and Experimental Condition (C), for each Subject (S) and Test Stimulus (T). Each column of diagrams corresponds to a different Test Stimulus (speech, SP; narrow-band noise of 150 Hz bandwidth, NBN150; and square wave, SQW; respectively.) The ordinate of each diagram shows the mean response level obtained by averaging over three test stimulus levels (L), and over two reference stimuli (R). On the abscissa, experimental conditions 1, 2, and 3 correspond to condition Q (quiet), P (partial-loudness), and O (overall-loudness), respectively. In each diagram, mean response levels obtained by adjusting the test stimulus are represented by + symbols and those by adjusting the reference stimulus are represented by □ symbols.

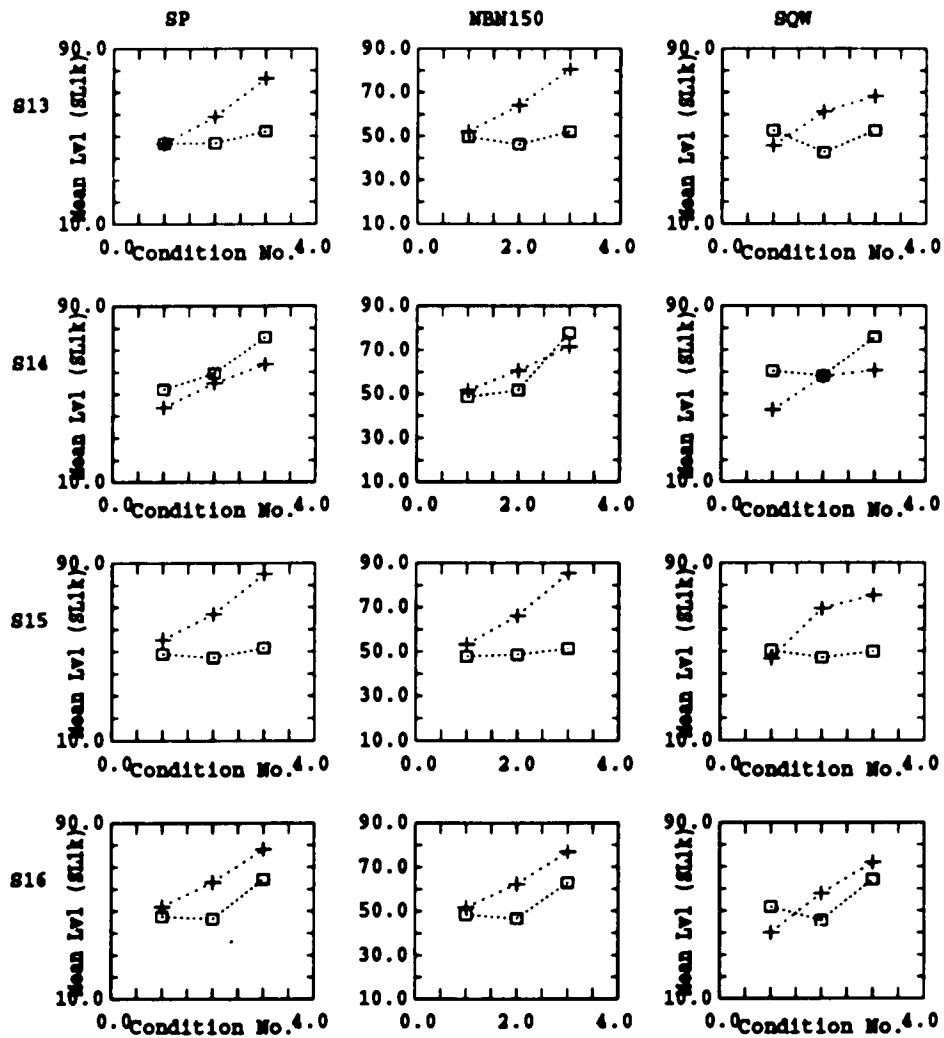


Figure A.14: Continuation of Previous Figure.

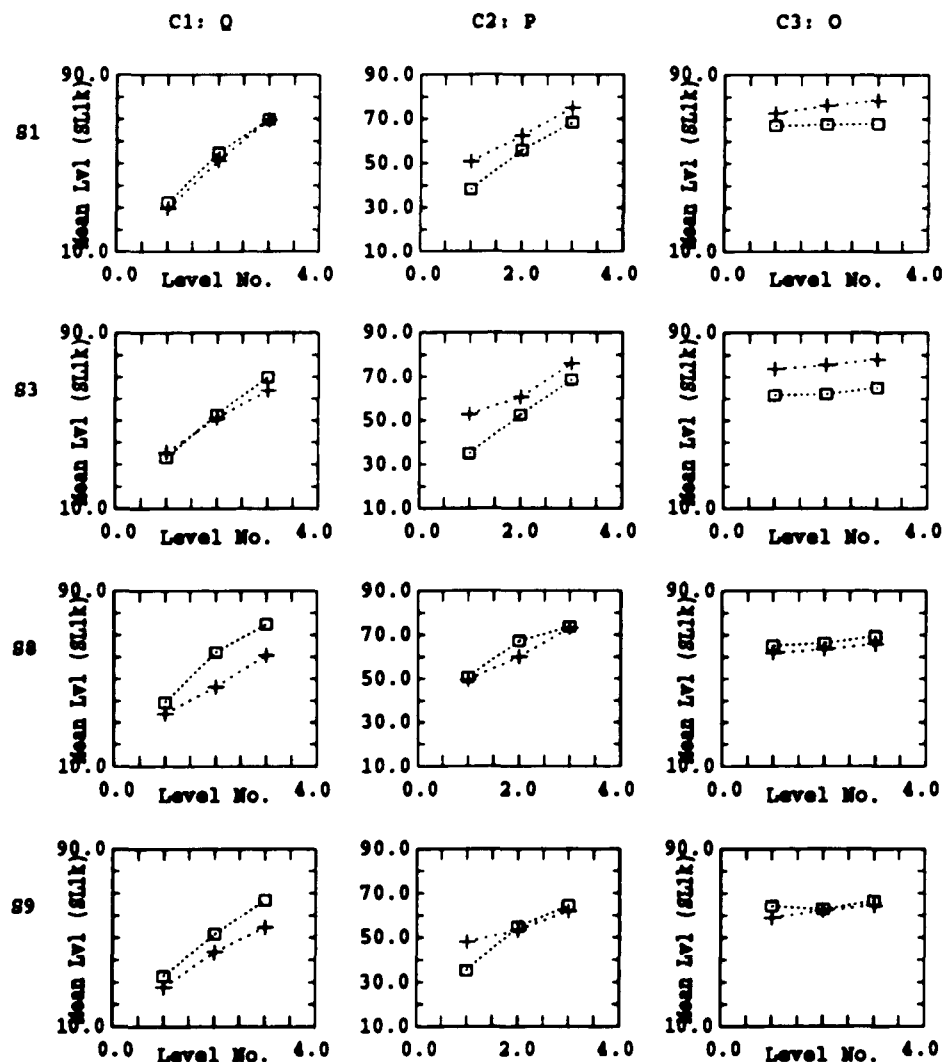


Figure A.15: Graphical Representation of LACS Interaction. Each diagram show the interaction between Adjustment Method (A) and Stimulus Level (L) for each Subject (S) and Experimental Condition (C). Each column of diagrams corresponds to a different Experimental Condition (Quiet, Q; Partial-loudness, P; and Overall-loudness, O; respectively.) The ordinate of each diagram shows the mean response level obtained by averaging over the test stimulus type (T) and over the reference stimuli (R). On the abscissa, the three stimulus levels, 1, 2, and 3, correspond to the low-, mid-, and high-levels, respectively. In each diagram, mean response levels obtained by adjusting the test stimulus are represented by + symbols and those obtained by adjusting the reference stimulus are represented by □ symbols.

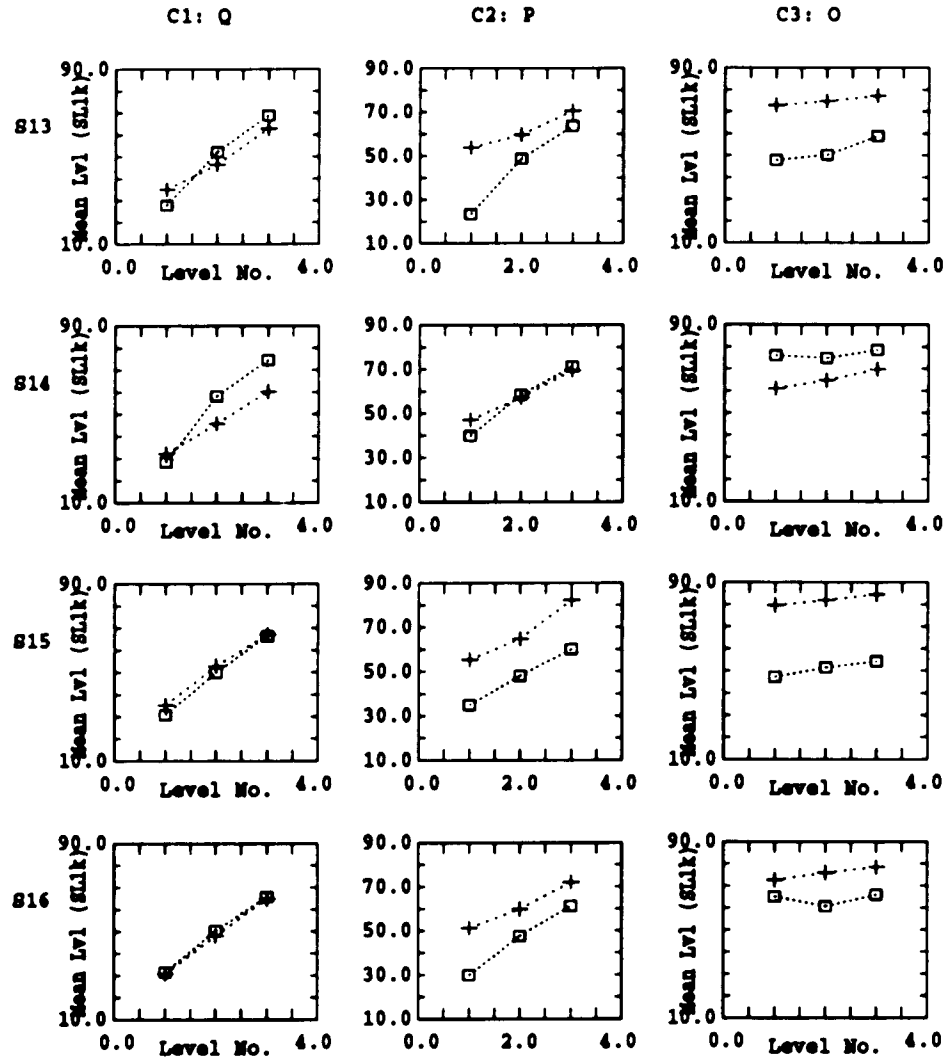


Figure A.16: Continuation of Previous Figure.

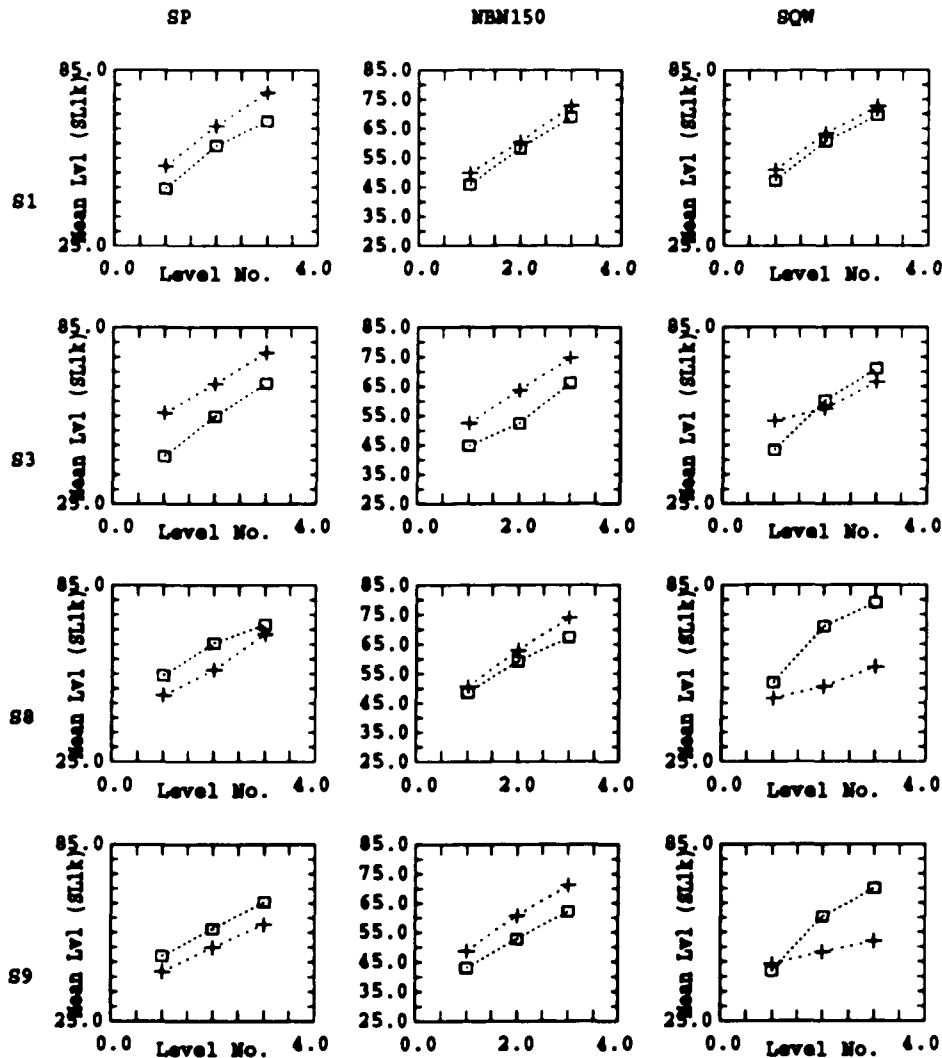


Figure A.17: Graphical Representation of LTAS Interaction. Each diagram show the interaction between Adjustment Methods (A) and Stimulus Level (L) for each Subject (S) and test stimulus (T). Each column of diagrams corresponds to a different Test Stimulus (speech, SP; narrow-band noise of 150 Hz bandwidth, NBN150; and square wave, SQW; respectively.) The ordinate of each diagram show the mean response level obtained by averaging over the experimental conditions (C), and over the reference stimuli (R). On the abscissa, the three stimulus levels, 1, 2, and 3, correspond to the low-, mid-, and high-levels, respectively. In each diagram, mean response levels obtained by adjusting the test stimulus are represented by + symbols and those obtained by adjusting the reference stimulus are represented by □ symbols.

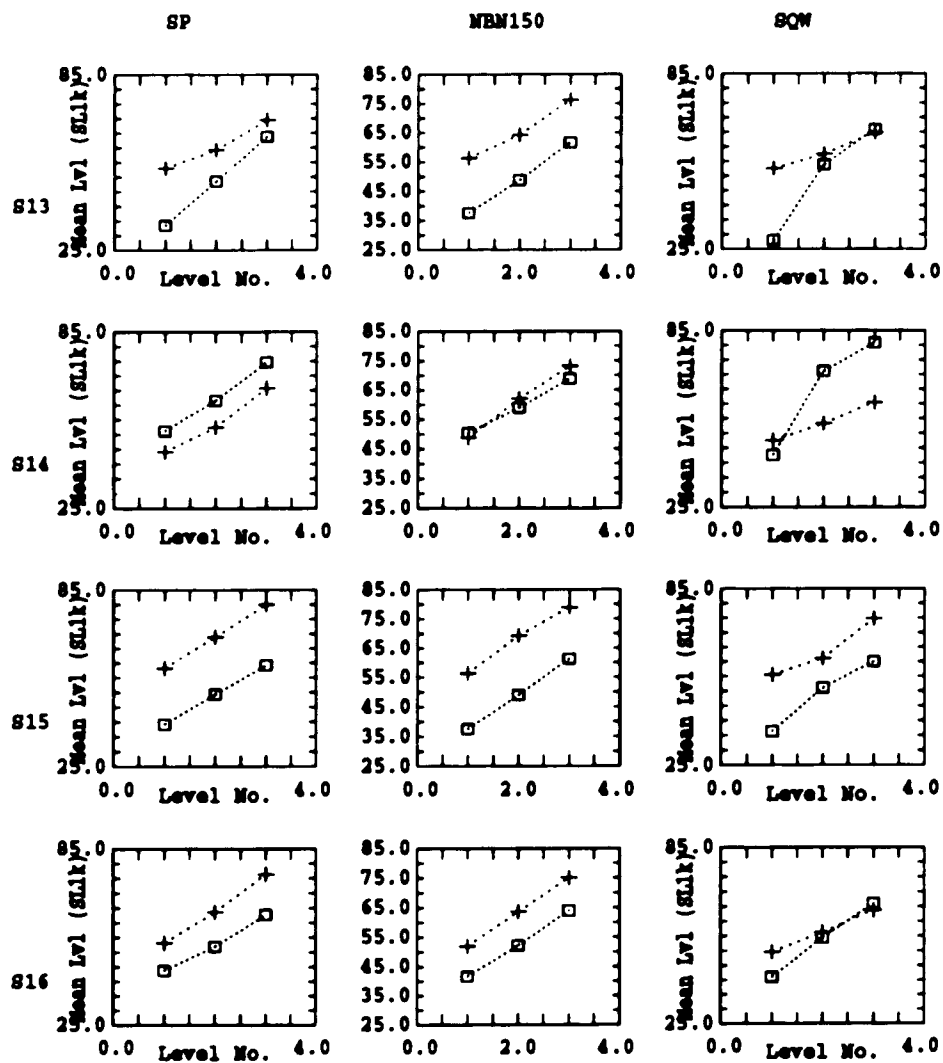


Figure A.18: Continuation of Previous Figure.

Loudness Growth Patterns

The analysis of variance showed that the choice of reference stimulus, pure tone or narrow-band noise, did not have a significant effect (or any significant interactions with other variables) on loudness matching. The loudness matching data were therefore averaged over the two reference stimulus types to obtain the loudness growth patterns for each subject. Figures A.19 through A.22 show the loudness growth patterns for each subject as a function of stimulus level (L) for each of the two adjustment methods (A). Separate loudness growth functions are shown for each test stimulus (T) and each experimental condition (C) for each subject (S). Each mean response level was obtained by averaging over the reference stimulus type (R). For each subject, there are nine plots corresponding to the nine combinations of three test stimuli by three experimental conditions. The three test stimuli correspond to the three columns of the plots and the three conditions to the three rows of the plots. In each plot, two symbols are used to represent the data obtained using the two adjustment methods, * for adjusting the test stimulus and o for adjusting the reference stimulus. Two subjects' data are presented in each figure. S1 and S3 are shown at the upper and lower halves of Fig. A.19 respectively; S8 and S9 of Fig. A.20; S13 and S14 of Fig. A.21; and S15 and S16 of Fig. A.22. By comparing the loudness growth patterns of the same Experimental Condition and of the same Test Stimulus across the subjects, i.e. one diagram per subject, one can observe that subjects showing odd patterns of loudness growth are those showing unusual interaction patterns, e.g. S8, S9, and S14. For example, comparing loudness patterns for speech in partial loudness condition across all subjects, for most subjects the loudness patterns obtained by the two adjustment methods are closely aligned except for subjects S8, S9, S14 and S15. Another obvious example,

is the loudness growth patterns for square wave in the overall loudness condition, where for most of the subjects the two patterns obtained by two adjustment methods are widely apart, that obtained by adjusting the reference stimulus is located to the left with slope near 0 while that obtained by adjusting the test stimulus is located to the right with slope near 1. For subjects S8, S9, and S14, the two patterns overlaps.

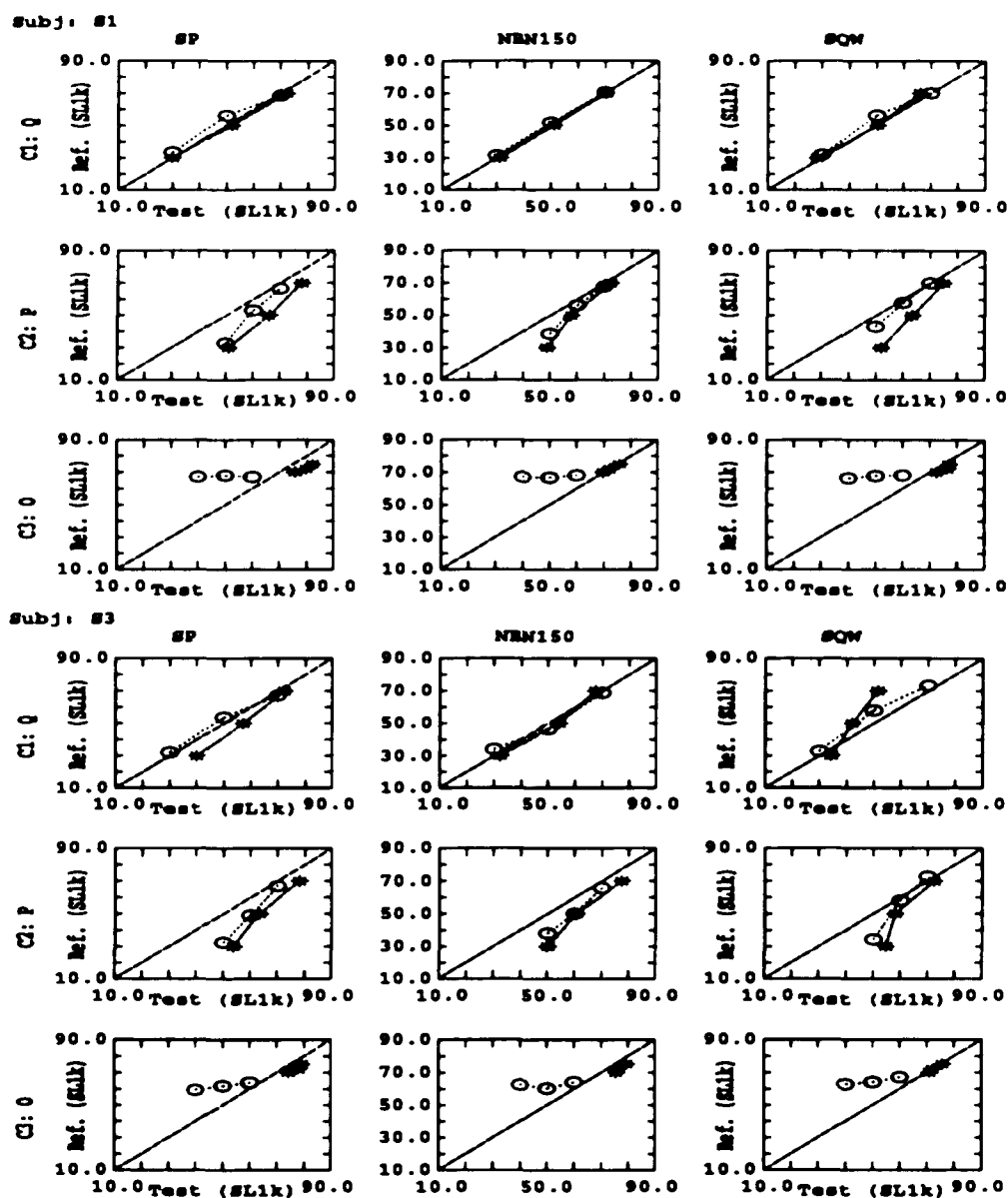


Figure A.19: Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S1 and S3. The ordinate of each diagram shows the level of the reference stimulus in sensation level re the subject's detection threshold of 1-kHz pure tone (SL1k). The abscissa shows the level of the test stimulus in SL1k. The open circles represent the loudness data obtained by adjusting the reference stimulus and the asterisks represent the loudness data obtained by adjusting the test stimulus.

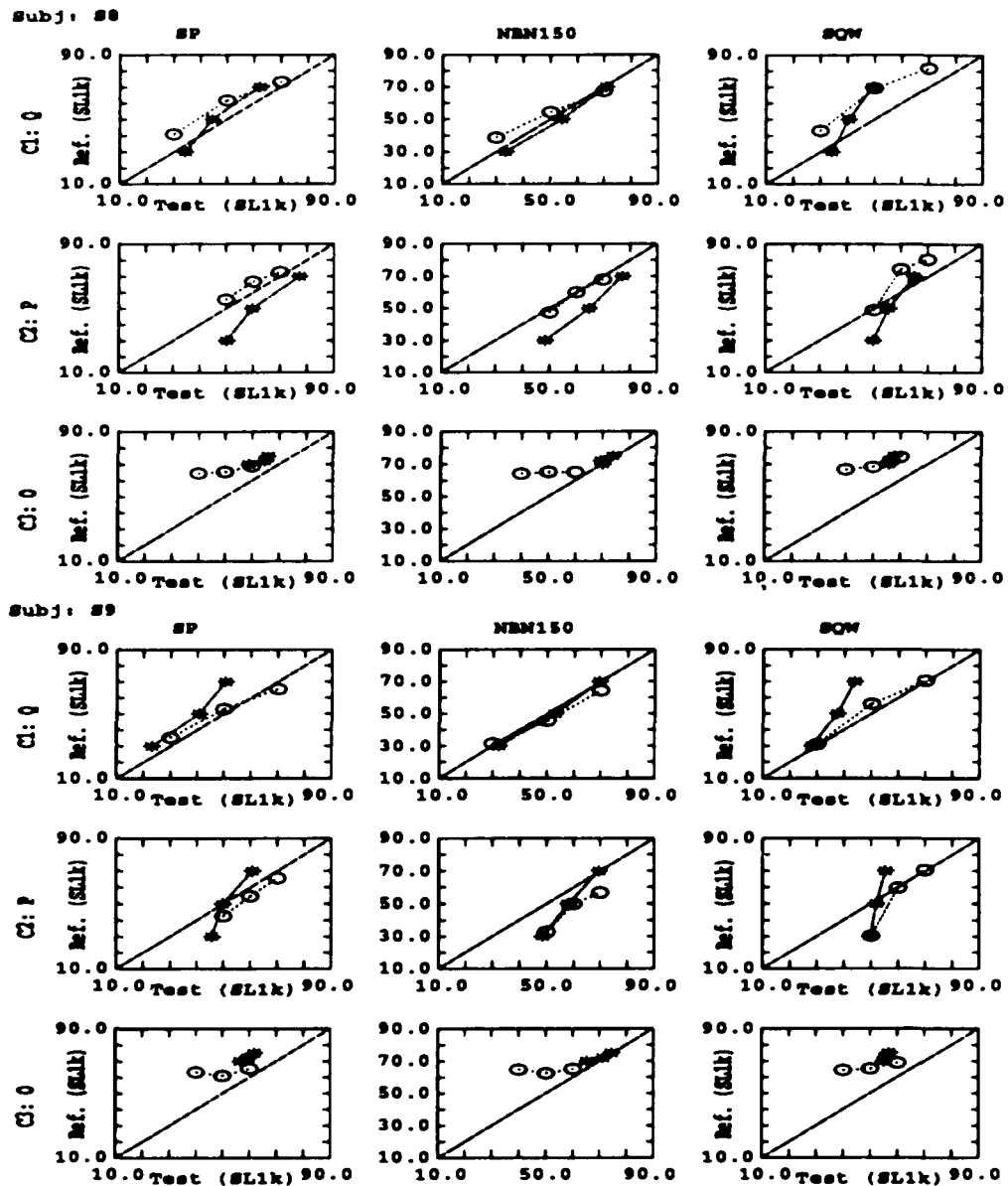
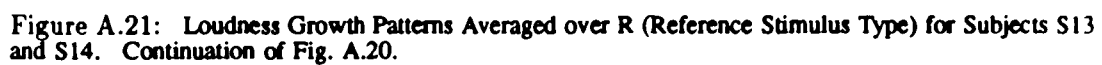


Figure A.20: Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S8 and S9. Continuation of Fig. A.19.



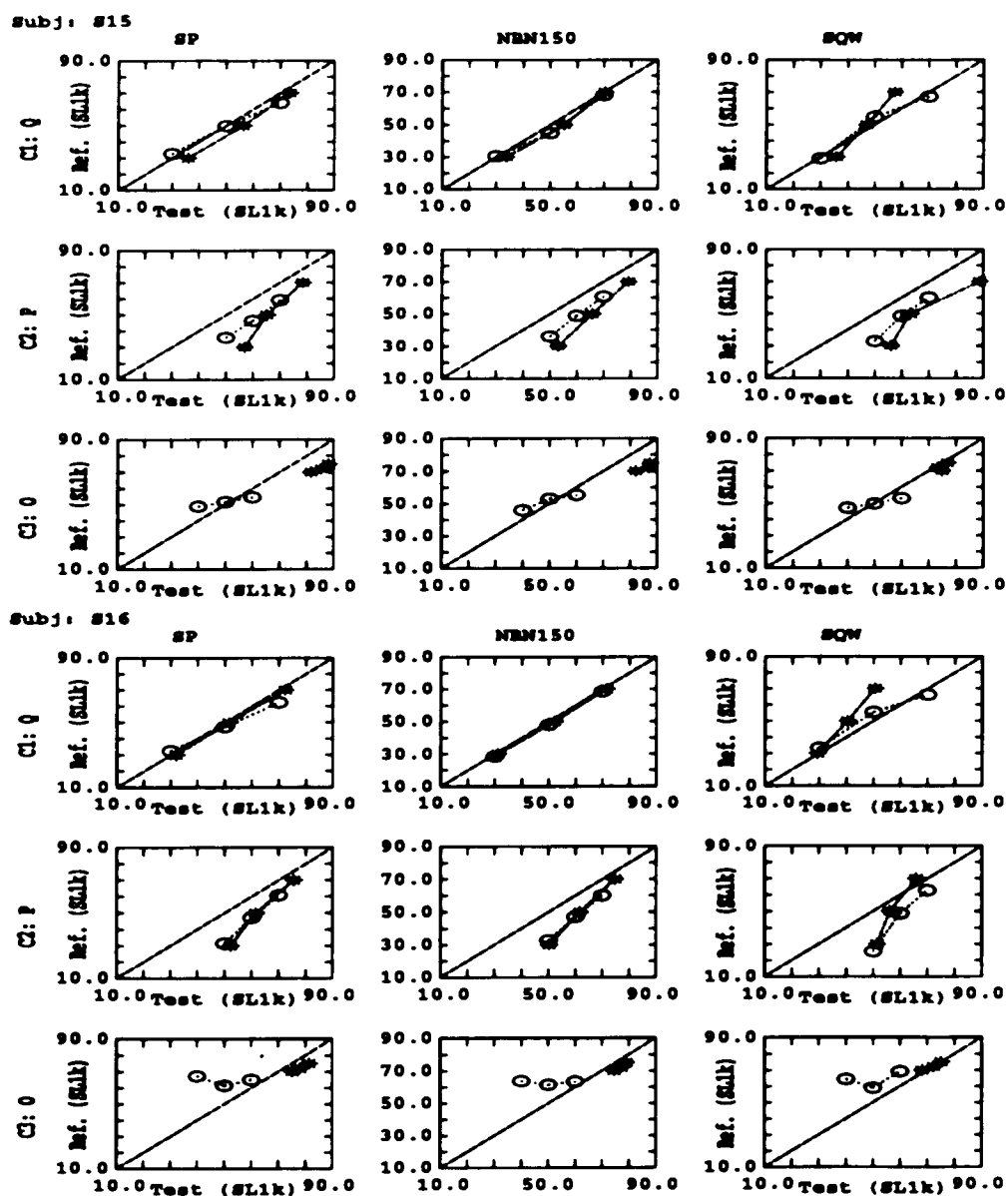


Figure A.22: Loudness Growth Patterns Averaged over R (Reference Stimulus Type) for Subjects S15 and S16. Continuation of Fig. A.21.

Bibliography

- [1] Abbas, P. H., and Sachs, M. B. (1976). "Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship," *J. Acoust. Soc. Am.* **59**, 112-122.
- [2] Allen, J. B. (1977). "Cochlear micromechanics - A mechanism for transforming mechanical to neural tuning within the cochlear," *J. Acoust. Soc. Am.* **62**, 930-939.
- [3] Allen, J. B. (1980). "Cochlear micromechanics - A physical model of transduction," *J. Acoust. Soc. Am.* **68**, 1660-1670.
- [4] Allen, J. B. (1983). "A hair cell model of neural response," in *Mechanics of Hearing*, edited by E. deBoer, M. A. Viergever (Delft University Press), pp. 193-202.
- [5] Allen, J. B. (1985). "Cochlear modeling," *IEEE ASSP Magazine*, January, 3.
- [6] Allen, J. B. (1989). "Is basilar membrane tuning the same as neural tuning - where do we stand?" in *Cochlear Mechanisms*, edited by J. P. Wilson and D. T. Kemp (Plenum Press, New York), pp. 453-463.
- [7] Allen, J. B., and Ghitza, O. (1990). "Analysis arrangement based on a model of human neural responses," *U. S. Patent*, 4905285.
- [8] Allen, J. B. (1991). (Personal communication).
- [9] ANSI (1980). ANSI S3.4-1980 / ASA 37-1980, "Procedure for the computation of loudness of noise," (American National Standards Institute, New York).
- [10] Arthur, R. M., Pfeiffer, R. R., and Suga, N. (1971). "Properties of 'two-tone inhibition' in primary auditory neurons," *J. Physiol. (London)* **212**, 593-609.
- [11] Beranek, L. L., Marshall, J. L., Cudworth, A. L., and Peterson, A. P. G. (1951). "Calculation and measurement of the loudness of sounds," *J. Acoust. Soc. Am.*, **23**, 261-269.
- [12] Carterette, E. C., and Friedman, M. P. (1978). *Handbook of Perception IV - Hearing*, (Academic, New York).
- [13] Cody, A. R. and Russell, I. J. (1987). "The responses of hair cells in the basal turn of the guinea pig cochlea to tones", *J. Physiol.* **383**, 551-569.

- [14] Miller, Jr., R. G. (1966). *Simultaneous Statistical Inference*, (McGraw-Hill Book Company, New York).
- [15] Delgutte, B. (1986). "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," presented at the NATO Advanced Research Workshop on *The Psychophysics of Speech Perception*, Utrecht, Netherlands, June 30-July 4, 1986.
- [16] Delgutte, B. (1989). "Physiological mechanisms of masking and intensity discrimination," in *Interactions between Neurophysiology and Psychoacoustics*, 81-101, 117th Meeting of the Acoustical Society of America.
- [17] Delgutte, B. (1990). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," *J. Acoust. Soc. Am.* **87**(2), 791-809.
- [18] Debout, P. (1959). "Observations of persistence of post-stimulus masking," *Acoustica* **9**, 353-358.
- [19] Evans, E. F. and Wilson, P. J. (1975). "Cochlear tuning properties: Concurrent basilar membrane and single nerve fiber measurement," *Science* **190**, 1218-1221.
- [20] Evans, E. F. and Palmer, A. R. (1980). "Relationship between the dynamic range of cochlear nerve fibers and their spontaneous activity," *Exp. Brain. Res.* **40**, 115-118.
- [21] Etymotic (1987). *Etymotic ER-2 Data sheet* (Oak Park, IL).
- [22] Fahey, P. F., and Allen, J. B. (1985). "Nonlinear phenomena as observed in the ear canal and at the auditory nerve," *J. Acoust. Soc. Am.* **77**, 599-612.
- [23] Fastl, H. (1977). "Loudness of running speech," *J. Audiol. Technique* **16**, 2-13.
- [24] Ferguson, G. A. (1981). *Statistical Analysis in Psychology and Education*, 5th edition, (McGraw-Hill, New York).
- [25] Fletcher, H., and Steinberg, J. C. (1924). "The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound," *Phys. Rev.* **24**, 306-317.
- [26] Fletcher, H., and Munson, W. A. (1933). "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.* **5**, 82-108.
- [27] Fletcher, H. (1935). "Newer concepts of the pitch, the loudness and the timbre of musical tones," *J. Franklin Instit.*, **220**, 405-429.
- [28] Fletcher, H., and Munson, W. A. (1937). "Relation between loudness and masking," *J. Acoust. Soc. Am.* **9**, 1-10.
- [29] Fletcher, H. (1938). "Loudness, masking and their relation to the hearing process and the problem of noise measurement," *J. Acoust. Soc. Am.* **9**(4), 275-293.
- [30] Fletcher, H. (1940). "Auditory patterns," *Review of Modern Physics* **12**, 47-65.

- [31] Fletcher, H. (1950). "A method of calculating hearing loss for speech from an audiogram," *J. Acoust. Soc. Am.* **22**(1), 1-5.
- [32] Fletcher, H. (1953). *Speech and Hearing in Communication*. (Van Nostrand Reinhold Company, New York,) pp. 153-175.
- [33] Gates, B. G. (1937). "Discussion before the meter and instrument section, 12th February, 1937," following "The performance of noise meters in terms of the primary standard," by B. G. Churcher and A. J. King, *J. Inst. Elec. Engrs.* **81**, 85-86.
- [34] Ghitza, O. (1985). "A measure of in-synchrony regions in the auditory nerve firing patterns as a basis for speech vocoding," presented by *International Conf. on Acoust., Speech and Signal Proc.*, ICASSP'85, Tampa, Florida, **2**, March, 505.
- [35] Ghitza, O. (1986). "Auditory nerve representation as a front-end for speech recognition in a noisy environment," AT&T Bell Laboratories Technical Memorandum.
- [36] Ghitza, O. (1987). "Robustness against noise: The role of timing-synchrony measurement," in *Proceeding IEEE 1987 International Conference on Acoustics, Speech, and Signal Processing*, Dallas, Texas.
- [37] Ghitza, O. (1988). "Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," *J. Phonetics* **16**(1), 109-123.
- [38] Goode, R. L., and Killion, M. C. (1987). "The middle ear from the standpoint of the surgeon and the acoustician," *J. Acoust. Soc. Am.* **81**.
- [39] Greenwood, D. D. (1961a). "Auditory masking and the critical band," *J. Acoust. Soc. Am.* **33**, 484-502.
- [40] Greenwood, D. D. (1961b). "Critical bandwidth and the frequency coordinates of the basilar membrane" *J. Acoust. Soc. Am.* **33**(10), 1344-1356.
- [41] Greenwood, D. D. (1990). "A cochlear frequency-position function for several species - 29 years later," *J. Acoust. Soc. Am.* **87**, 2592-2605.
- [42] Guild, (1932). "Correlations of histologic observations and the acuity of hearing," *Acta Oto-Laryng.* **17**, 207.
- [43] Guinan, J. J., and Peak, W. T. (1967). "Middle ear characteristics of anesthetized cats," *J. Acoust. Soc. Am.* **41**, 1237-1261.
- [44] Hall, J. L. (1977a). "Spatial differentiation as an auditory 'second filter': Assessment on a nonlinear model of the basilar membrane," *J. Acoust. Soc. Am.* **61**, 520-524.
- [45] Hall, J. L. (1977b). "Two-tone suppression in a nonlinear model of the basilar membrane," *J. Acoust. Soc. Am.* **61**, 802-810.
- [46] Hall, J. L. (1981). "Observations on a nonlinear model for motion of the basilar membrane," in *Hearing Research and Theory*, edited by J. V. Tobias and, E. D. Schubert (Academic Press, New York), pp. 1-61.

- [47] Harrison, J. M. and Howe, M. E. (1974). "Anatomy of the descending auditory system (mammalian)," in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin), Vol. 5/1, pp. 363-388.
- [48] Hellman, R. and Zwislocki, J. (1961). "Some factors affecting the estimation of loudness," *J. Acoust. Soc. Am.* **33**(5), 687-694.
- [49] Hellman, R. P. (1970). "Effect of noise bandwidth on the loudness of a 1000-Hz tone," *J. Acoust. Soc. Am.* **48**, 500-504.
- [50] Hellman, R. P. (1976). "Growth of loudness at 1000 and 3000 Hz," *J. Acoust. Soc. Am.* **60**(3), 672-679.
- [51] Hellman, R. P. (1978). "Dependence of loudness growth on skirts of excitation patterns," *J. Acoust. Soc. Am.* **63**(4), 1114-1119.
- [52] Hellman, R. P. (1991). (Personal communication).
- [53] ISO (1975). **ISO 532-1975(E)**. "Acoustics - method for calculating loudness level," *International Standard*.
- [54] Johnson, D. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J. Acoust. Soc. Am.* **68**, 1115-1122.
- [55] Johnstone, B. M., Patuzzi, R. and Yates, G. K. (1986). "Basilar membrane measurements and the travelling wave," *Hearing Res.* **22**, 147-156.
- [56] Kimberley, B. (1991). (Personal communication).
- [57] Khanna, S. M. and Leonard, D. G. B. (1982). "Basilar membrane tuning in the cat cochlear," *Science* **215**, 305-306.
- [58] Knauss, H. P. (1937). "An empirical formula for the loudness of a 1000-cycle tone," *J. Acoust. Soc. Am.* **9**, 45-46.
- [59] Kryter, K. D. (1959). "Scaling human reactions to the sound from aircraft," *J. Acoust. Soc. Am.* **31**, 1415-1429.
- [60] Liberman, M. C. (1978). "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.* **63**(2), 442-455.
- [61] Liberman, M. C. (1982a). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," *J. Acoust. Soc. Am.* **72**(5), 1441-1449.
- [62] Liberman, M. C. (1982b). "Single-neuron labelling in the cat auditory nerve," *Science* **216**, 1239-1241.
- [63] Liff, H. J., and Goldstein, M. H., Jr. (1970). "Peripheral inhibition in auditory fibers in the frog," *J. Acoust. Soc. Am.* **47**, 1538-1547.
- [64] Littler, T. S. (1965). *The Physics of the Ear*. Pergmon Press, New York, pp.99-104.

- [65] Mintz, F. and Tyzzer, F. G., (1952). "A loudness chart for octave-band data on complex sounds," *J. Acoust. Soc. Am.* **24**(1), 80-82.
- [66] Möller, A. R. (1983). *Auditory Physiology*, (Academic Press, New York), 40.
- [67] Moskowitz, H. R. (1968). "Scales of intensity for single and compound tastes," *PhD Dissertation* (Harvard University).
- [68] Nedzelnsky, V. (1980). "Sound pressures in the basal turn of the cat cochlea," *J. Acoust. Soc. Am.* **68**, 1676-1689.
- [69] Neely, S. T., and Kim, D. O. (1983). "An active cochlear model showing sharp tuning and high sensitivity," *J. Acoust. Soc. Am.* **79**, 1472-1480.
- [70] Neff, W. D. (1961). "Neural mechanisms of auditory discrimination," in *Sensory Communication*, edited by W. A. Rosenblith (The M.I.T. Press, Mass. and John Wiley & Sons, Inc., New York), pp. 259-278.
- [71] Quitzsch, G. (1955). *Acustica* **5**; *Akust. Beih.* **1**, 49-66.
- [72] Pang, X. D. and Peake, W. T. (1986). "How do contractions of the stapedius muscle alter the acoustic properties of the ear?" in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely and A. Tubis (Springer, Berlin), pp. 36-43.
- [73] Peterson, A. P. G. and Beranek, L. L. (1954). *Handbook of Noise Measurement* (General Radio Company, Cambridge, Mass.).
- [74] Patuzzi, R., Sellick, P. M., and Johnstone, B. M. (1984). "The modulation of the sensitivity of the mammalian cochlea by low-frequency tones. III. Basilar membrane motion," *Hearing Res.* **13**, 19-27.
- [75] Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing* (Academic Press, London), Second Edition.
- [76] Plomp, R. and Bouman, M. A. (1959). "Relation between hearing threshold and duration of tone pulses," *J. Acoust. Soc. Am.*, **31**(6), 749.
- [77] Puria, S. and Allen, J. B. (1991a). "A parametric study of cochlear input impedance," *J. Acoust. Soc. Am.*, **89**, 287-309.
- [78] Puria, S. (1991b). "A theory of cochlear input impedance and middle ear parameter estimation," *PhD Dissertation* (City University of New York).
- [79] Puria, S. (1992). (Personal communication).
- [80] Rhode, W. S. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* **49**, 1218-1231.
- [81] Rhode, W. S. (1978). "Some observations on cochlear mechanics," *J. Acoust. Soc. Am.* **64**, 158-176.

- [82] Robinson, D. W., and Dadson, R. S. (1956). "A Re-determination of the equal-loudness relations for pure tones," *Br. J. Appl. Phys.* **7**, 166-181.
- [83] Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlear. I. Input-output functions, tuning curves, and response phases," *J. Acoust. Soc. Am.* **80**(5), 1364-1374.
- [84] Robles, L., Ruggero, M. A., and Rich, N. C. (1986a). "Mössbauer measurements of the mechanical response to single-tone and two-tone stimuli at the base of the chinchilla cochlea," In *Peripheral Auditory Mechanisms* (eds J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely and A. Tubis), pp. 121-128. Springer, Berlin.
- [85] Rosowski, J. J. (1991). "The effects of external- and middle-ear filtering on auditory threshold and noise-induced hearing loss," *J. Acoust. Soc. Am.* **90**, 124-135.
- [86] Sachs, M. B., and Kiang, N. Y. S. (1968). "Two-tone inhibition in auditory-nerve fibers," *J. Acoust. Soc. Am.* **43**, 1120-1128.
- [87] Scharf, B. (1959). "Critical bands and the loudness of complex sounds near threshold," *J. Acoust. Soc. Am.* **31**, 365-370.
- [88] Scharf, B. (1961). "Loudness summation under masking," *J. Acoust. Soc. Am.* **33**, 503.
- [89] Scharf, B. (1978). "Loudness," in *Handbook of Perception IV - Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 187-242.
- [90] Sellick, P. M., and Russell, I. J. (1979). "Two-tone suppression in cochlear hair cells," *Hearing Res.* **1**, 227-236.
- [91] Sellick, P. M., Patuzzi, R. and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131-141.
- [92] Sellick, P. M., Patuzzi, R. and Johnstone, B. M. (1983). "Comparison between the tuning properties of inner hair cells and basilar membrane motion," *Hearing Res.* **72**, 131-141.
- [93] Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoust. Soc. Am.* **56**(6), 1848-1861.
- [94] Sheeley, E. C. and Bilger, R. C. (1964). "Temporal integration as a function of frequency," *J. Acoust. Soc. Am.* **36**, 1850-1857.
- [95] Sivian, L. J. and White, S. D. (1933). "Minimum Audible Sound Fields," *J. Acoust. Soc. Am.* **4**, 288.
- [96] Small, A. M. (1962). "Loudness as a function of Signal Duration," *J. Acoust. Soc. Am.* **34**(4), 513-514.
- [97] Steinberg, J. C. (1937). "Positions of stimulation in the cochlear by pure tones," *J. Acoust. Soc. Am.* **8**, 176-180.

- [98] Steinberg, J. C. and Gardner, M. B. (1937). "The dependence of hearing impairment on sound intensity," *J. Acoust. Soc. Am.* **9**, 11.
- [99] Stevens, J. C., and Marks, L. E. (1965). "Cross-modality matching of brightness and loudness," *Proc. Nat. Acad. Sci.* **54**, 407-411.
- [100] Stevens, J. C., and Stevens, S. S. (1963). "Brightness function: Effects of adaptation," *J. Opt. Soc. Amer.* **53**, 375-385.
- [101] Stevens, S. S. (1936). "A scale for the measurement of a psychological magnitude: Loudness," *Psych. Rev.* **43**, 405-416.
- [102] Stevens, S. S. (1956). "Calculation of the loudness of complex noise," *J. Acoust. Soc. Am.* **28**(5), 807-832.
- [103] Stevens, S. S. (1957a). "Calculating loudness," *Noise Control* **3**(5), 11-22.
- [104] Stevens, S. S. (1957b). "On the psychophysical law," *Psychological Review* **64**, 153-181.
- [105] Stevens, S. S. (1961). "Procedure for calculating loudness: Mark VI," *J. Acoust. Soc. Am.* **33**(11), 1577-1585.
- [106] Stevens, S. S. (1966). "Matching functions between loudness and ten other continua," *Perception Psychophys.* **1**, 5-8.
- [107] Stevens, S. S., and Greenbaum, H. (1966). "Regression effect in psychophysical judgement," *Perception Psychophys.* **1**, 439-446.
- [108] Stevens, S. S. (1971). "Perceived level of noise by Mark VII and decibels (E)," *J. Acoust. Soc. Am.* **51**(2.2), 575-601.
- [109] Stevens, S. S., and Davis, H. (1963). *Hearing: Its Psychology and Physiology*, (the American Institute of physics, Inc., 2nd., New York.)
- [110] Tukey, J. W. (1949). "Comparing individual means in the analysis of variance," *Biometrics* **5**, 99-114.
- [111] Viemeister, N. F. (1988). "Intensity coding and the dynamic range problem," *Hearing Res.* **34**, 267-274.
- [112] von Békésy, G. (1928). "Zur Theorie des Hörens; Die Schwingungsform der BasilarMembran," *Physik. Zmits.*, **29**, 793-810.
- [113] von Békésy, G. (1960). *Experiments in Hearing*. (McGraw-Hill, New York).
- [114] Warr, W. B. (1978). "The olivocochlear bundle: its origins and terminations in the cat," in *Evoked Electrical Activity in the Auditory Nervous System*, edited by R. F. Naunton and C. Fernandez (Academic Press, New York and London.) pp. 43-63.
- [115] Warr, W. B. and Guinan, J. J. (1979). "Efferent innervation of the organ of Corti: two separate systems," *Brain Res.* **173**, 152-155.

- [116] Watson, C. S. and Gengel, R. W. (1968). "Time-intensity trading relations as a function of signal frequency," *76th Meeting of the Acoustical Society of America*, Paper A7.
- [117] Wever, E.G. (1949). *Theory of Hearing*, (Wiley, New York).
- [118] Whitfield, I. C. (1978). "The Neural Code," in *Handbook of Perception IV - Hearing*, edited by E.C. Carterette and M.P. Friedman (Academic, New York,) pp.163-183.
- [119] Young, E.D., and Barta, P. E. (1986). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," *J. Acoust. Soc. Am.* **79**(2), 426-442.
- [120] Zwicker, E. and Feldtkeller, R. (1955). "Über die Lautstärke von gleichförmigen Geräuschen," *Acustica* **5**, 303-316.
- [121] Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical band width in loudness summation," *J. Acoust. Soc. Am.* **29**(5), 548-557.
- [122] Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," *J. Acoust. Soc. Am.* **33**(2), 248.
- [123] Zwicker, E. and Scharf, B. (1965). "A model of loudness summation," *Psychological Review* **72**(1), 3-26.
- [124] Zwicker, E. (1977). "Procedure for calculating loudness of temporally variable sounds," *J. Acoust. Soc. Am.* **62**(3), 675-682.
- [125] Zwicker, E., Fastl, H., and Dallmayr, C. (1984). "BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B," *Acustica*. **55**, 63-67.
- [126] Zwicker, E. (1985a). "What is a meaningful value for quantifying noise reduction?" *Proceedings for Inter-noise '85*, 47-56.
- [127] Zwicker, E., Deuter, K., and Peisl, W. (1985b) "Loudness meters based on ISO 532B with large dynamic range," *Proceedings for Inter-Noise '85*, 1119-1122.
- [128] Zwicker, E. (1987). "Procedure for calculating partially masked loudness based on ISO 532 B," *Proceedings for Inter-noise '87*, 1021-1024.
- [129] Zwicker, E., and Zwicker, T. (1991). "Audio engineering and psychoacoustics: Matching signals to the final receiver, the human auditory system," *J. Audio. Eng. Soc.* **39**(3), 115-126.
- [130] Zwislocki, J. J. (1980). "Five decades of research on cochlear mechanics," *J. Acoust. Soc. Am.* **67**, 1679-1685.
- [131] Zwislocki, J. J. and Kletsy, E. J. (1980). "Micromechanics in the theory of cochlear mechanics," *Hearing Res.* **2**, 505-512.